MSc thesis

Computer Science

Modeling emotions in dialogue generation

Essi Salmenkivi

April 10, 2020

Faculty of Science

University of Helsinki

Supervisor(s)

Prof. Hannu Toivonen

Examiner(s)

Prof. Hannu Toivonen, Dr. Mark. Granroth-Wilding

Contact information

P. O. Box 68 (Pietari Kalmin katu 5)
00014 University of Helsinki,Finland

Email address: info@cs.helsinki.fi
URL: http://www.cs.helsinki.fi/

HELSINGIN YLIOPISTO – HELSINGFORS UNIVERSITET – UNIVERSITY OF HELSINKI

Tiedekunta — Fakultet — Faculty	Koulutusohjelma — Utbildningsprogram — Study programme
Faculty of Science	Computer Science

Tekijä — Författare — Author

Essi Salmenkivi

Työn nimi — Arbetets titel — Title

Modeling emotions in dialogue generation

Ohjaajat — Handledare — Supervisors

Prof. Hannu Toivonen

Työn laji — Arbetets art — Level	Aika — Datum — Month and year	Sivumäärä — Sidoantal — Number of pages
MSc thesis	April 10, 2020	74 pages, 7 appendice pages

Tiivistelmä — Referat — Abstract

This work introduces a system for generating radio play scripts. Generating dramatic dialogue presents unique challenges in language generation. In addition to fluency of language, dramatic text should exhibit plot and characters' affective stances to each other and events. Character relationships and affect may be expressed beneath the surface level of everyday conversation topics. In the affect-driven dialogue generation system introduced by this thesis, characters have goals, relationships and a three-dimensional model of mood which influences their behaviour. Given conflicting goals, characters will navigate the web of conversation, making choices that influence others to accept their goal while simultaneously trying to maintain the relationship to others. Characters react emotionally to each others' speech acts and express their own affective state in how they speak. The system separates the form of a sentence from its content, allowing the system to generate a wide range of coherent, dramatic conversations by combining affect-expressing sentence templates with goal-expressing content. Because content and form are independent from each other, only a finite number of sentence templates need to be prepared to generate conversations about any content.

ACM Computing Classification System (CCS)
Computing methodologies → Artificial intelligence → Natural language processing → Natural language generation

Avainsanat — Nyckelord — Keywords

natural language generation, drama, dialogue generation

Säilytyspaikka — Förvaringsställe — Where deposited

Helsinki University Library

Muita tietoja — övriga uppgifter — Additional information

Algorithms specialisation line

Contents

1 Introduction

We are living in the golden age of artificial intelligence or AI. Computational problem solving and decision making are being utilized in a almost every field imaginable from trading in the stock exchange to recycling garbage. Computational creativity, or employing computers to build creative artifacts such as pictures, poetry or music, is sometimes considered the final frontier of AI [8].

In the past decades, artificial intelligence has enjoyed success in many problems where the definition of success is fairly straightforward and computable, such as chess (remove the opponent's pieces from the board according to simple rules) and image recognition (classify sets of pixels). In art, the values are more difficult to quantify in terms of numbers. What is the mathematical formula for humour or beauty?

Creativity is a complex concept that is so inherent to the human experience that it can be difficult to define what is and isn't creative. Creative processes produce creative *artifacts*, something that is new and valuable in some way. A widely-used model of creativity breaks the discussion into four perspectives, the four P's [24]. The four facets of creativity are *product* (the artifact that can be consumed via reading or hearing), *person* (the subject driving the creative process), *process* (the act of creating) and *press* (the social setting which receives the product). This view of creativity emphasizes the nature of creativity (and art) as not only a subjective but an interpersonal phenomenon that cannot be evaluated just by looking at any of the four facets on its own.

Wiggins [31] frames computational creativity as a search for valuable artifacts in a universe of possible artifacts. In order to find valuable artifacts, the builder of the creative system must define evaluation criteria to filter valuable artifacts from uninteresting ones. Such criteria depend on the context in which the generated product is to be consumed.

The project to build the generative system described in this thesis started as a collaboration project between University of Helsinki's Discovery research group and YLE, the Finnish broadcasting company. The goal of the project was to explore the use of artificial intelligence in drama by creating a computer system that outputs Finnish language radio play scripts that can be produced by human directors and actors to record a radio play for airing on radio.

In building the radio play generation system, the goal is to generate radio play scripts that are coherent. Natural language generation based on simple statistical predicting the next word has created interesting and amusing artifacts but is expensive, both in terms of computation and input data. The quality of artifacts generated in this manner is also very erratic. In the radio play generation system, freedom of generation is restricted by rules that ensure the generated products are coherent.

Drama is fundamentally about characters and their relationships [15], even if the surface-level plot is about other topics, even banal ones. The aim of the radio play system is to model characters with inner lives including goals, emotions and relationships with one another. These characters are given a situation in which their goals are at odds with each other. The central purpose of the experiment is to test what kind of variations can be created within a story restricted to this structure, and whether this setup is enough to create interesting tension in a conversation that is at the surface talking about everyday topics.

The core hypothesis that the radio play system tests is that the topics of conversation can be separated from the structures of expression, which reflect the emotional states of the speakers. A conversation about the same topic can be expressed in many ways depending on the emotional states and relationships between the speakers. For example, a married couple at a dinner table can ask each other to pass the salt, while the *way* this request is made may express tension, resentment or love between them. In drama, this is called subtext [9]. Subtext is the story happening beneath the explicit words said out loud, the reason why things are said.

The characters or agents in the radio play generation system must therefore have a representation of their goals, internal states and relationships with each other. Unlike many narrative generation systems, the radio play generation system does not output events or states, but lines of dialogue. This presents unique challenges to the system, as it must find a way to parameterize lines of dialogue and map them onto emotional states.

2 Related work

Research in narrative generation, modeling emotions and dialogue generation is plentiful. Work combining all three, however, is scarce.

Gebhard introduces ALMA - A Layered Model of Affect [13] as an emotional reasoning model for a 'virtual human' - a character that interacts via dialog and facial expressions with other such characters and humans. Gebhard defines emotion, mood and personality as three temporally different layers of the same phenomenon, emotions being short-term affects, mood medium-term and personality long-term.

Gebhard uses an updated version of the OCC (Ortony, Clore and Collins) model [20] of emotions [29] to build an affect space consisting of three dimensions: pleasure, arousal and dominance. Gebhard's virtual humans are affected by events which are classified as direct or indirect (caused by a character or the world at large), desirable or non-desirable (positive or negative effect on a character) and praiseworthy or not praiseworthy (positive or negative action by a character). Events are appraised in different ways depending on each character's personality. The classification of events finally outputs an emotion which affects the agent's mood. When no events happen, the mood slowly normalizes. Gebhard's system produces a flexible model of fluctuation of emotions, but it does not produce story or plot on its own. Instead, it is designed to be used with an event-generating system such as the VirtualHuman [13].

Berov introduces a motivation driven story generator [4]. In Berov's system, characters with a modelled personality and mood are used as agents who internally reason and output actions that they are motivated towards in each emotional state. Berov aims to simulate the folk story "The Little Red Hen" using a multi-agent simulation where agents have affect (modeled using the Big Five personality model) and beliefs, desires and intentions. Agents' personalities determine what intentions they select and how they respond to other agents' actions. Character decisions are a product of a cycle that takes in events and evaluates them using the character's emotional appraisal. In Berov's system, the course of events is highly dependent on the agents' personalities: with certain traits, agents will develop no desires and no plot emerges.

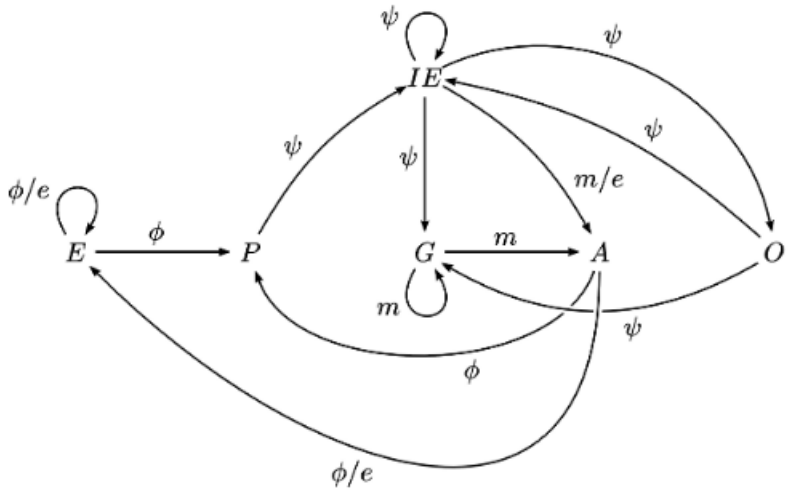
The belief-desire-intention model [23] (BDI) is an important part of Berov's architecture. The BDI model assumes an environment where agents have imperfect or only local infor-

mation about the state of the environment, not perfect global information. An agent's information is called its beliefs. The environment is dynamic and can evolve in multiple different ways at any point in time, creating a decision tree branching out into possible states, which an agent can choose from. In order to make such choices, agents have an evaluation function for each possible state. Positively evaluated states make up the agent's desires. Some desires may be mutually incompatible, so the agent must finally use a deliberation function to pick out a path in the decision tree that maximizes expected utility. The states along that path are ones that the agent is committed to pursuing, or intentions.

Affect-driven story generation is explored by Pérez y Pérez [21] in the MEXICA system. The MEXICA system models characters' emotions as directed emotional links between characters. An emotional link is affected by actions, for example having one's life saved elicits gratitude and strengthens a positive emotional link. On the other hand, emotional links also influence the actions taken by characters (and thus the story) by acting as preconditions for certain actions. For example, the action *kill* has a strong negative emotional link as a precondition for being taken.

A story-generation system focusing on causal relations between events is introduced by Swartjes & Theune [30]. This system generates consecutive events using the GTN (General Transition Network), a graph in which nodes are events and (directed) edges are different types of causal relations between events (figure 2.1). Any path that follows the edges will generate a logically coherent chain of events, or fabula.

Figure 2.1: Graph for generating causally connected story elements [30]



Fabula is a different concept than plot. Whereas plot includes all the events narrated in

the story, the fabula also includes events that are causally connected to plot events but aren't explicitly narrated. In other words, the set of plot events is a subset of fabula events in the story model. Swartjes & Theune do not implement a method for selecting the most interesting subset of events to create a story.

The measure that is maximized when deciding on a subset of events to narrate is called *tellability* [5]. In addition to ensuring that the events are interesting to the reader, tellability must guarantee that events are connected in a logical fashion and the resulting plot is understandable and leaves no problematic inconsistencies or unexplained events.

3 Dialogue and conversation analysis

The format of a radio play script is dialogue. Dialogue consists of speech acts taken by one or more characters. Scripts may also include stage directions, which are descriptions of events or actions by the characters, written in third person. Stage directions might look like "Airplane takes off in distance" or "Man gets out of car." The radio play generation system focuses on dialogue generation, leaving stage directions out.For the sake of simplicity, stage instructions are not considered in the radio play generation system, and the only text generated is the dialogue itself.

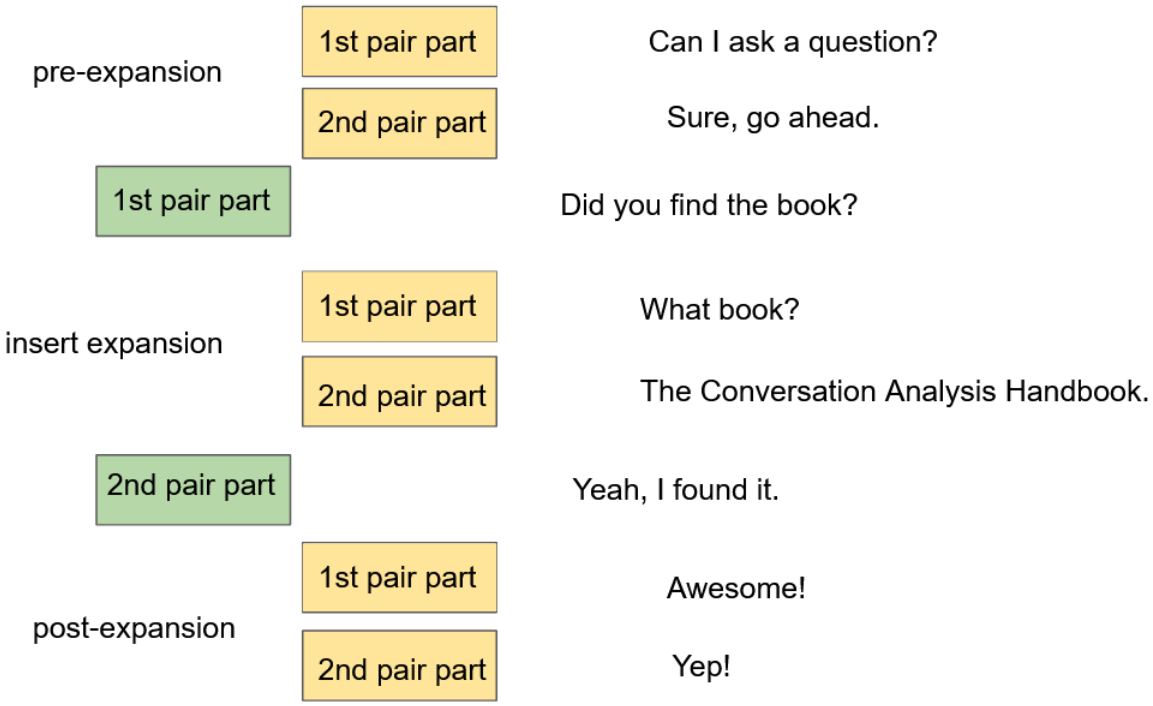
In order to build a system that generates structurally coherent dialogue, we look at what tools linguistics provide for understanding dialogue. Specifically, we look at the subfield of sociolinguistics called conversation analysis.

Conversation analysis is the study of the structure of natural, everyday conversation. In conversation analysis, the atomic building block of conversation is a *turn* [27]. A turn is a unit of speech taken by one speaker. One of the fundamental questions of conversation analysis is the 'turn-taking' problem — who talks next at any given time. Conversation analysis aims to understand how turns get organized, what makes larger structures coherent and how meaning arises from subtle elements across turns.

Individual turns are arranged into *adjacency pairs*. Conversation analysis assumes that all turns are part of a pair of two turns, which are in a call-response relationship to each other. Seemingly unpaired turns, like a question without an answer are considered to be implicitly paired with a special turn called the empty turn, or $\emptyset$.

Adjacency pairs are further arranged into *sequences*. A sequence consists at minimum of one adjacency pair, but the core adjacency pair may also be expanded via pre-expansions (before the first pair part), insert expansions (between the first and second pair part) and post-expansions (after the second pair part). Figure 3.1 illustrates a sequence with a core adjacency pair, one pre-expansion, one insert expansion and one post-expansion. Expansions are subordinate to the core adjacency pair, meaning that they cannot exist without the core adjacency pair, but the core adjacency pair does not need the expansions. Expansions are also independent of each other: any one of them may be removed and the rest of the dialogue is still understandable.

Figure 3.1: Sequence with expansions.



The role of pre-expansions (before main pair) is often to orient the discussion towards the main topic. Insert expansions (between main first and second pair part) and post-expansions (after the main adjacency pair) may ask for clarifications or more information about the main pair. Post-expansions may also serve as a ”post-mortem”, a summary of the topic that has been discussed.

In conversation analysis, turns and adjacency pairs are understood in terms of the *actions* they relate to. The action perspective enables us to classify turns by the relation to the action. A turn may inform, request, agree, disagree, complain, soothe, question, answer, contest and so on. Adjacency pairs and sequences may be given an action label and the turns within it are given more specific labels. For example, a question sequence may consist of a question first pair part and an answer second pair part. Labeling sequences and turns is not uniform across the field of conversation analysis. New labeling systems will be created to serve the purposes of each study.

Conversation analysis also sees turns and sequences as functional in relation to the larger context of the social interaction that speech is part of: the *project* [25]. In conversation analysis, project is the concept that organizes and motivates sequences. Two different definitions of the project exist in CA: there is the micro-project that describes the in-

tention of a single turn, such as ”getting permission” or ”informing about absence”. The macro-project refers to larger units which may be broken down into activities. A macro-project may span over several sequences. An example of a macro-project is the ”medical project”, which can be broken down into activities like ”description of symptoms”, ”asking additional questions” and ”giving diagnosis”.

The radio play generation system uses these conversation analysis concepts to model lines of dialogue and the ways they are organized. The system takes a limited set of action categories which are used to build the dialogue. The action categories are designed in such a way that projects can be realized via any action type without the meaning changing in a way that would affect the story.

4 Overview of the radio play generation system

In this section, we give an overview of the structure of the radio play generation system. In section 4.1, we definitively formulate the goals of the system on different levels of abstraction. In section 4.2 we examine the architectural choices of the system, including the separation of the form of turns (section 4.2.2) and their content (section 4.2.3). Finally, in section 4.3, we examine how the system models the entities in the story world.

4.1 Coherence

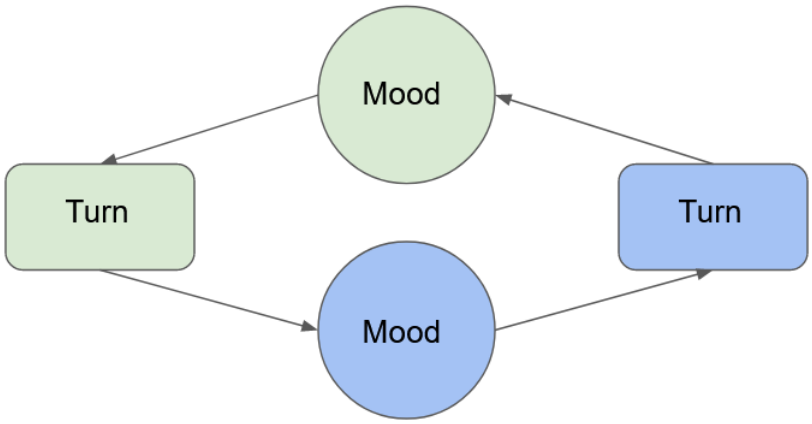
The goal of the radio play generation project is to build a system that can produce coherent scripts that could be acted out by human voice actors and aired on radio to the entertainment of listeners. The system should allow as much variance in the outputted scripts as possible given the restraint of coherence on three levels: structural, factual and affective.

Structural coherence requires that sentences are intact and follow the grammatical rules of the Finnish language. It also requires that sentences that follow each other do so in a realistic manner, i.e. the question ”how are you doing” should be answered with ”thanks, I’m doing well” or ”none of your business” but not ”my grandmother’s vase” or ”no I won’t”. Structural coherence is taken care by the hierarchical turn generation and turn type system, explained in section 4.2.

Statistical language models are one solution to generating text that upholds structural coherence. This has been done with models ranging from simple Markov models to recent models such as BERT [10] and GPT-2 [22]. However, even these state-of-the-art models that consist of hundreds of millions of parameters fail to exhibit factual or affective coherence. This project aims to explore ways in which factual and especially affective coherence can be supported with a rule-based framework that is as light as possible, and one that could be later augmented using a statistical language model to generate language.

Factual coherence aims to ensure that the topics of the characters’ conversation across

Figure 4.1: Simplified dialogue generation cycle: one character's turn affects the other's mood and vice versa.



larger structures than sentences are logical. Statements about the world should be consistent, for example if a vase is blue in the first scene it should not be red in the second scene. Factual coherence is the responsibility of the story world, described in section 4.3.

Finally, the main focus of the thesis is affective or emotional coherence. The radio play generation system models its characters as agents that have emotional states (moods) that govern their behaviour. If the dialogue generated by the system has affective coherence, it should express the moods of the character in a believable, consistent way in each turn.

The affective system should also be dynamic, in other words, the characters' moods should change logically according to what is said to them. A threat should be reacted to with fear or anger, not joy or relief. Figure 4.1 illustrates the affective system as a cycle, each character reacting to what the other says. The way the system models characters' moods and emotions is covered in detail in section 5.

In addition to expressing characters' moods, affective coherence also includes expressing characters' relationships to each other. In the end, drama is about people and their relationships. Although the plot may discuss everyday events, the dialogue should express *subtext*, the tension between the characters. Subtext is the underlying implicit argument beneath the explicit topics discussed openly.

The challenge of a rule-based system is the trade-off between writing a lot of rules and range of artifacts generated. The aim in this work is to find a minimal set of rules that allows the maximum variation while maintaining coherence. We can also view variation from the same three perspectives as coherence: structural, factual and affective. Structural variation ensures that the flow of conversation does not follow the same pattern every time.

Factual variation means that topics of conversation vary, and affective variation means that the characters can exhibit a wide range of affective states.

In this phase of development, the plot generation system aims to explore how much variety can be created in radio play scripts by modifying only the affective level, while events stay the same. For the sake of simplicity, the plot is confined into two scenes and the number of characters is limited to two. This means that factual variation is not implemented. In later development, the content (characters and events) can also be generated more freely.

In the restricted two-scene format, scene one features character A calling character B and informing them that their mutual relative has died. In scene two, the characters meet at the relative's apartment and argue over inheritance, specifically one item that they both want.

Two possible design philosophies can be considered for maintenance of coherence. We can build hard coherence, where coherence is the first priority and incoherence cannot be generated. The second option is soft coherence, where generation doesn't consider coherence, but focuses on other values (such as entertainment or variation) first. In this case, coherence is evaluated after generation and artifacts that don't satisfy coherence are filtered out. This evaluation can be done automatically within the program or by humans.

Many computational creativity systems employ soft coherence, often human-evaluated. The content generated using language models often varies greatly in quality and the high-quality artifact that is selected for publication is usually an outlier in terms of coherence and value. For the layman observer, this final filtering step may be opaque, which may cause the general public to view the creative capability of computational creative systems as more advanced than it actually is.

For the sake of transparency, the radio play generation system is built on the principle of hard coherence. Although restricting the generated artifacts more in order to maintain hard coherence may diminish the range of possible artifacts, the benefit is that the artifact that is finally selected to air on radio will be close to an average artifact instead of an outlier. Human evaluation to find the best artifact to produce and air on radio will still be employed, but the goal is that any and every artifact produced will be presentable.

4.2 Structure

The dialogue generation system should take as input the pre-decided content of the story (events) and return a script that is one of the structurally coherent ways to tell the events. The radio play system utilizes concepts from conversation analysis to model dialogue and control its generation. In this section, we examine how the concepts from conversation analysis are used to serve the needs of this dialogue generation system, one of which is the ability to produce scripts of varying form. The length of the outputted radio play script should not be fixed, but should vary from a minimalist, straightforward script where each scene is performed in a few lines to potentially infinitely long back and forth argument between the characters.

4.2.1 Adjacency pairs and extensions

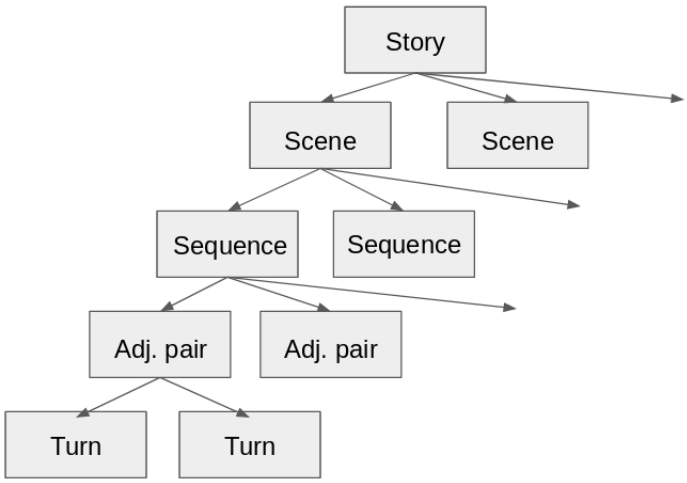
The fundamental structural principle in conversation analysis is that every turn, or line of dialogue, belong to an adjacency pair consisting of a call and a response. This principle proves useful in the radio play system as well.

Basic structure: adjacency pairs In order to hold structural coherence, the generation rules must include a way to make sure questions are answered and that they are answered only once. The solution to this problem is to use an adjacency pair structure, inspired by principles of conversation analysis as presented in section 3. The adjacency pair design dictates that all lines uttered by a character are part of an adjacency pair: in other words, each line is either a first pair part (question, proposition, etc.) or a reply to such. Generating turns as part of an adjacency pair makes sure that each question gets answered and that it gets answered only once.

The story could be generated linearly, adjacency pair after adjacency pair. In this architecture, each bottom-level item (sentence) in the story must take into account all of the information in the system, in other words, a sentence is a function of all the previous sentences so far.

The other design option is a hierarchical tree structure, where each node only needs the information from its parent node. In this design, nodes can be generated independently of nodes in other branches. The top node contains the fixed information about the story world and its inhabitants. The leaf nodes are individual sentences or *turns*. Turns are

Figure 4.2: Basic structure of radio play dialogue.



combined into adjacency pairs consisting of a first pair part (eg. question, command) and second pair part (eg. answer, compliance, refusal).

Figure 4.2 depicts the four-layered architecture of the radio play generation system. The root node of the conceptual plot tree is the story. The story generates a number of scenes, each of which contains one major story event. Scenes consist of sequences. Sequences always have a purpose to the character starting it: characters start sequences to further their goals, which are encoded in the form of projects (discussed later in section 4.2.3). For now, we can think of projects as sequences the character wants to start.

A sequence consists of a main adjacency pair and zero or more expansion adjacency pairs. The turns in an adjacency pair are always spoken by different characters. The character speaking a given turn is called the *speaker* of the turn, while the other character is the *listener*. In an adjacency pair, we can also say that the speaker of the first pair part is the speaker, and the speaker of the second pair part is the listener.

Expansions The radio play system is structured so that it generates radio play script recursively, as scenes can contain any number of sequences and sequences can contain any number of adjacency pairs. The benefit of this design is that topics of conversation can be processed within a sequence and its recursive children. The child adjacency pairs, or expansions, are subordinate to their parent adjacency pairs. For example, table 4.1 shows a main adjacency pair appended with a nested insert expansion adjacency pair. The resulting sequence can be read with the expansion, but it also makes sense if the insert expansion is left out. The expansion cannot exist alone, without the main adjacency pair.

Pekka	I want the vase
Kalle	Why haven't you told me you want the vase?
Pekka	You heard me
Kalle	No! Not the vase

Table 4.1: Expansions are subordinate to parent nodes.

Using this recursive generation structure, a minimal radio play generated by the system includes two scenes, each of which contains one sequence. Each sequence contains one adjacency pair whose first pair part claims the scene's event and the second pair part accepts it. At its simplest, with no expansions or additional sequences, the generated radio play could look like this:

Scene 1

Pekka: Grandmother has died.
Kalle: Oh no.

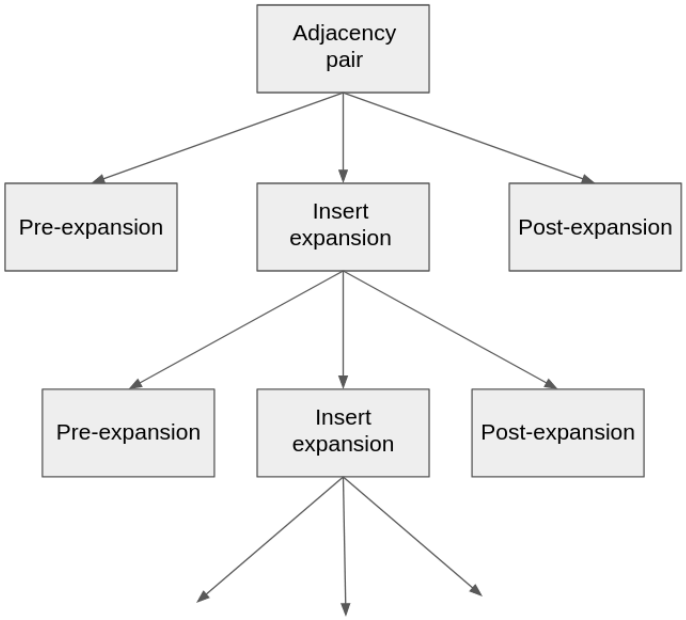
Scene 2

Kalle: I want the vase.
Pekka: Sure, you can have the vase.

In a non-minimal script, scenes will contain multiple sequences. Additional sequences which are independent from the main events can be greetings, goodbyes and discussions of weather, for example. Another way to generate variety in the structure of radio plays is to lengthen sequences with additional adjacency pairs: expansions. Sequences may be expanded with additional adjacency pairs before, between or after the main adjacency pair. Expansions may further be expanded with expansions of their own, recursively, as illustrated in figure 4.3.

In addition to adding independent sequences, the radio play is lengthened by *refusal* sequences. A character's goal is to receive an accepting reply (as in the example above) to their statements and proposals, and when they receive it, they stop talking about the subject. However, if the listener does not accept, the speaker will try to push their agenda again and again, creating additional sequences. In other words, refusal fails to satisfy the goal that prompted the speaker to start the sequence in the first place.

Figure 4.3: Adjacency pairs are expanded by other adjacency pairs, recursively.



First pair part	Can I take the vase?	
Refusal	No, you are not taking the vase.	No way.
Agreement	Sure, take the vase.	Go ahead.

Table 4.2: The two groups of second pair parts and examples of them.

The sequence-adjacency pair-turn architecture governs the structure of the radio play scripts on a high level. To see how structural coherence is maintained inside individual adjacency pairs, in other words, how the system makes sure questions and replies stay relevant to each other, we will look at turn types.

4.2.2 Turn types

The dialogue generation system should be able to generate a linguistically rich variety of different sentences. However, since the main events in the story are fixed, the system needs a way to generate variety within a sentence while expressing the same event. The variety should not constrain the combinatory creativeness of the system: in other words, if new expressions are added to the system, they should be possible to use in combination with more than one existing expression. Alternative expressions of the same sentiment are called turn types.

Turn type	Realization project 1	Realization project 2
Type 1	I want to take the vase.	I want to eat the pizza.
Type 2	Do you mind if I take the vase?	Do you mind if I eat the pizza?
Type 3	I really need to take the vase.	I really need to eat the pizza.

Table 4.3: Examples of turn types

For example, character A requesting to take grandma's vase could be expressed by saying in a neutral manner: "Can I have the vase?", politely: "If you don't mind, I'll be taking the vase" or more dominantly: "I'm taking the vase and you have nothing to say about it". Conversely, B's refusal of giving the vase to A could be expressed neutrally with "No, the vase is mine", politely with "You aren't taking the vase, are you?" or aggressively: "Over my dead body". In order to maintain freedom of combinatory creativeness, these alternative expressions of the same intent must be chosen in such a way that any of the request turn types can be answered by any of the refusal turn types.

If each request turn type can take one of two specific replies, four request turn types with two possible replies each gives $4 \times 2 = 8$ different adjacency pairs with 12 different turn types. More generally, this design can generate $\frac{n}{3} \times 2$ adjacency pairs for n enumerated turn types. On the other hand, if any request turn can be combined with any reply turn, 8 different turn types outputs $6 \times 6 = 36$ variations of adjacency pairs. In the generalized form, this design gives $(\frac{n}{2})^2$ adjacency pairs for n turn types.

In order to keep turn types reusable for different conversation topics, the radio play system separates the form of a turn, turn type, from its content, the *project* (see section 4.2.3). Turn types can be thought of as blank templates of sentences into which words from the project are inserted, while projects are the subjects, verbs and objects that fill the template. Table 4.3 shows examples of three turn types, Type 1, Type 2 and Type 3 built with two different projects (columns).

Turn type categories While being able to combine any turn types freely would allow superior linguistic variance, unrestricted combining makes it difficult for the system to maintain coherence. In order to manage the meaning of the dialogue, turns are split into groups: first pair parts into statement, proposal and surprise and second pair parts into acceptance and refusal. Within a group, meaning should be roughly preserved. This means that no matter what sentence is picked from the group of statements regarding topic X, the conversation can continue with the assumption that X has been communicated. Figure

Figure 4.4: Turn types are divided by their function.

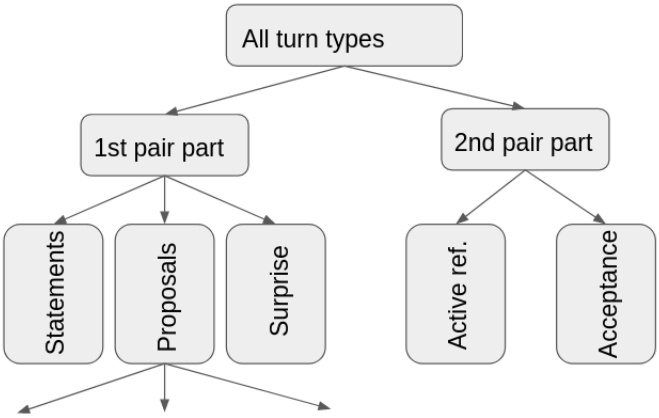
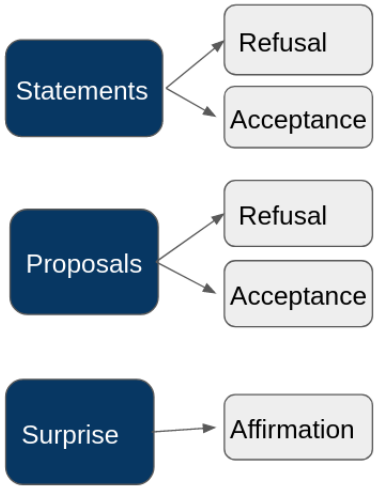


Figure 4.5: First pair part categories (dark) and their possible second pair part categories.



4.4 illustrates the categories or groups of turn types.

Statement turn types and proposals cannot be answered using the same sentences. Although at a base level, all refusals and acceptance could be expressed as ”No” and ”Yes”, respectively, any more specific, richer expressions will be incompatible with either type of first pair part. For example, refusals to the *statement* ”Grandmother has died” like ”That can’t be true”, ”You must be mistaken”, ”No, she hasn’t” can’t very fluently be used in response to a *proposal* such as ”I’m taking the vase” or ”Give me the vase.” Therefore, each first pair part category has its own set of refusal and acceptance turns. Figure 4.5 illustrates the relationships between turn type categories.

Surprise turns are a third type of way to start a sequence. According to Clark et al. [7] [12],

the brain is a system that minimizes entropy or prediction errors. We use this view to model the decision making process of the characters in the radio play system. When a character hears something surprising, in other words something they didn't predict correctly, their reaction will be to update their model of the world to improve future predictions. Surprise turns aim to express this internalization of new information, taking forms such as "Are you serious?" and "Why am I only hearing about this now?".

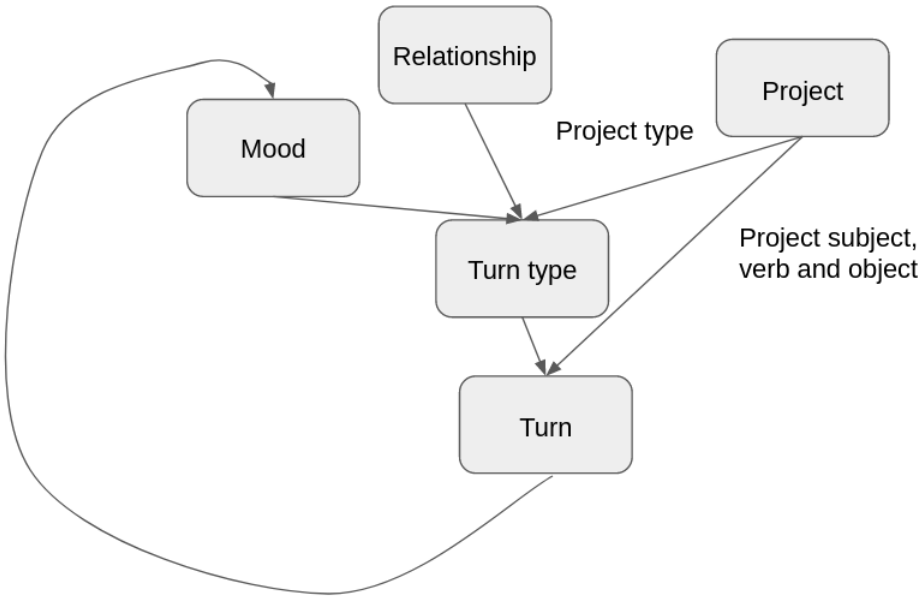
Modeling surprise as prediction error gives the system a way to motivate characters to start new topics that naturally relate to the original, main topic. In future development, characters could have a longer conversation about the topic that elicited surprise, bringing richness to the whole script. In the case of a relative's death, the surprised character could ask questions to update his model of the world and events leading to the death, such as "When did grandmother die?" or "How did grandmother die?"

Categorizing second pair parts Second pair parts come in two varieties: *acceptance* and *refusal*. The type of second pair part depends on whether the listening character considers the statement valid (in the case of statements) or agrees with the proposal (in the case of proposals). This is determined by considering whether the statement aligns with the character's knowledge of the world, or if the proposal aligns with their goals.

Surprise first pair parts are an exception to this and can only be answered in one way. Since there are only two characters and surprise is a reaction to a statement or proposal, the response to a surprise is always made by the character who originally spoke the surprise-eliciting statement or proposal, we can determine that surprise will always be responded to by re-affirming the original statement or proposal. We will call responses to surprise first pair parts *affirmations*.

Turn types are thus a system of structuring characters' expressions into groups with rules that allow for maximum variability while maintaining structural coherence. The next problem to solve is what different turn types are available in a given situation, and how the program decides which one to use. Now, since structural coherence is taken care of, the objective becomes affective coherence. Maintaining affective coherence requires a model of affect that governs how characters' affective states evolve and show in speech. This model should function in two ways: first, turn types should be selected in a way that expresses characters' mental states or moods. Second, turn types used by a character should affect the other character's mood. Figure 4.6 shows the generation of turns from mood as a cyclical process.

Figure 4.6: Overall structure of the turn formation. Every turn affects the other character's mood and relationship with the other character, which in turn affect the turn type they choose next.



As a result of this cyclical mood-turn-mood design, the way turn types are differentiated within a turn type group is by their association with character mood. For example, the proposal group of turn types includes turns with very different affective flavors, such as the polite request "Can I please have the vase?", the command "Give me the vase" and the affirmation "I need to take the vase". The turn type used in a turn is selected from the appropriate group (statement, proposal or surprise for first pair parts, acceptance or refusal for second pair parts) using affective reasoning based on the character's mood and relationship to the other character.

Similarly, second pair part acceptance turns vary in affective flavor. The request for the vase can be accepted politely ("Here you go, it's yours"), passive-aggressively ("Fine, take it") or sulkily ("I was afraid you would take it"). The logic of this turn type selection based on mood is described in section 6. First, we will examine how turn content or projects are handled and how moods and relationships are modelled.

Unaligned turn type, unaligned project
Can I have the vase?
Thank you for telling me that grandmother has died.
Unaligned turn type, aligned project
Can I have the vase?
Thank you for telling me that you take the vase.
Aligned turn type, unaligned project
Can I have the vase?
Sure, grandmother can die.
Aligned turn type, aligned project
Can I have the vase?
Sure, take the vase.

Table 4.4: Coherence requires that both turn type and project align in an adjacency pair.

4.2.3 Project

A character-focused story is driven by the intentions of its characters [4]. In a radio play script, lines of dialogue should always have a purpose in the context of a story, either driving the plot forward of revealing information about the characters to the listener. Dialogue should not be generated for the sake of conversation, but feel meaningful to the overall story. On the other hand, questions and intents introduced in the story should be carried out to a satisfying conclusion, not left open.

The solution to making sure each line of dialogue carries importance to the plot is the *project*. The project in the radio play generation system is an abstraction of a statement or event. Each turn in the script expresses exactly one project. The project gives the turn its content, while the turn type gives it form. While turn types make sure the generated conversation is structurally coherent, projects uphold factual coherence. If the first and second pair part of an adjacency pair are selected from compatible turn type categories and both exhibit the same project, the resulting adjacency pair will be both structurally and factually coherent. Table 4.4 illustrates an adjacency pair with conflicting turn types and projects.

The turns in table 4.4 have been selected to illustrate the problems of unaligned turn types. Many turn types are also project-agnostic, as shown in table 4.5. These turn types are all second pair parts and are only coherent when the first pair part does express the

Why are you telling me this now?
That's not true.
You're wrong.
Well that's just great.

Table 4.5: Examples of turn types that don't use project information.

project information explicitly, giving the second pair part context in which it is interpreted. Project-agnostic turns employ relative pronouns such as *this* and *that*, which refer to the first pair part, as well as sentences that as a whole refer to something that has already been said, such as the sentence *You're wrong [about that]*.

Many of these project-agnostic turn types could be coherently used with more than one type of first pair part: for example. "That's not true" could be a reply to either the *statement* "Grandmother has died" or the *proposal* "I'm taking the vase", although the latter is arguably less fluent. However, each acceptance and refusal turn type is set to only belong to one category (statement, proposal or surprise), even if the wording could work for multiple categories.

Even if a turn doesn't use project information to form its sentence, each turn keeps track of its project, as it may need to pass the project information on to child (expansion) sequences.

High-level projects determine the course of the story. Since characters are active agents in the radio play generation system, projects are not independent entities that move the story regardless of the characters. Instead, each project is owned by a character. When characters speak, they generate turns using the content information from their own projects. In the first scene, one character has the project "inform other character that relative has died". In the second scene, both characters have the project "gain dead relative's vase". In other words, projects describe what characters want to achieve in the conversation.

Internalized projects, therefore, serve as goals: a simplified version of the belief-desire-intention model [23]. Characters act (initiate new sequences in the conversation) based on their goals. Projects are blueprints for speech acts (turns) and are considered completed when a turn constructed using the project is answered with an accepting second pair part. When characters have no more unresolved projects, the scene ends.

As turn types that make up first pair parts are divided into statement, proposal and surprise, each project contains information about which turn types it needs to be used

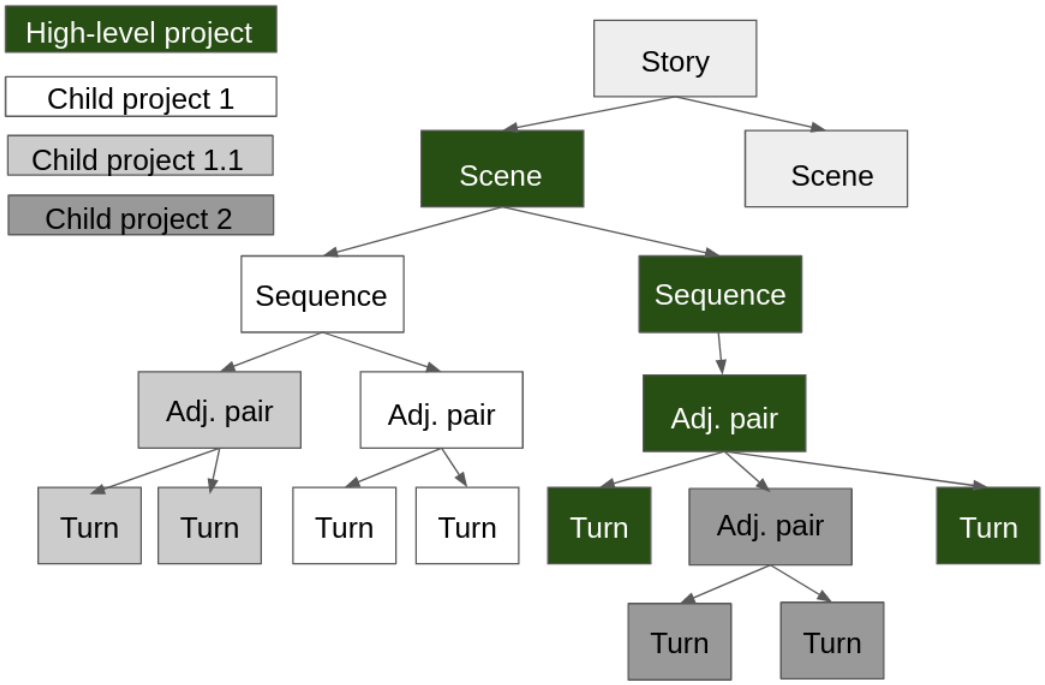
with. The high-level project ”inform other character that relative has died” has the type *statement* and the project ”gain dead relative’s vase” corresponds with the turn category *proposal*. Any turn types from the correct category can be combined with the project to form sentences.

In the radio play generation system, high-level projects are prescribed at the beginning of generation. In future development, it would be interesting to generate even top-level projects automatically. If the restriction of factual coherence was lifted, this could be done easily by generating projects using random verbs and nouns. However, if we want to maintain factual coherence instead of generating absurd dialogues, selecting coherent verbs and nouns for projects would require a large knowledge base giving the program information about causal relations of events or, if we allow an error rate, a language model that selects words that co-occur in a sentence with a high probability.

Projects appear on each level of the story hierarchy, top-level projects being the major plot points and lower-level projects minor conversation topics like the weather. Every bottom-level node, i.e. turn, in the generation tree, corresponds to exactly one project. Top-level projects propagate down from the story level to individual turns, making sure that the main plot points are realized in the dialogue. As a sequence generates its set of adjacency pairs, it first makes sure that the project passed to the sequence is given to one of the adjacency pairs. Then the sequence populates the rest of the adjacency pairs with new projects taken from a pool of neutral projects such as speaking about the weather, or creates new projects from the attributes of concepts mentioned in the main project. Figure 4.7 illustrates how top-level projects are propagated downwards in the radio play tree. Both turns of an adjacency pair always express the same project.

Top-level projects can yield child projects that are expressed in other branches of the radio play tree. The one example of a child project is the surprise project. When a character is surprised by another character’s turn, they gain a surprise project, which is a causally inferred project related to the project causing the surprise. In the radio play’s topics, characters know that *death* is caused by *sickness* and *taking* something is caused (or preceded) by *wanting* it. Table 4.6 illustrates a statement turn and the surprise turn yielded by it. Having a separate surprise project with new words adds richness to the language, but is not necessary for fluency. If the system has no causal information to generate a surprise project, it can also use the original project to generate surprise turns. Like with generating more projects, generating logical surprise projects automatically requires a knowledge base: each added project added needs a surprise project

Figure 4.7: Propagation of project from higher levels to leaf nodes.



Statement	Grandmother has died.
Surprise with surprise project	I didn't know grandmother had gotten sick.
Surprise with statement project	I didn't know grandmother was dead.
Proposal	I'm taking the vase.
Surprise with surprise project	I didn't know you wanted the vase.
Surprise with proposal project	I didn't know you are taking the vase.

Table 4.6: Surprise turns can be realized with a surprise project or the parent project.

that expresses an event that logically precedes or causes it.

Child projects can also expand on the topic of its parent project, focusing on some detail of sub-part of it. Generating sub-projects is exemplified in table 4.7. Generating sub-projects is based on attributes of previously mentioned objects or characters. Any attribute of an object or character may be used to expand the conversation, and this can be done recursively as long as the new attribute also has attributes. The generation of child projects in this way requires that the program has a knowledge base that includes descriptive attributes for its objects.

Now that we have discussed what projects are used for, we can finally precisely define what

Sentence	target	attributes
What vase are you talking about?	vase	owner: mother
Mother's vase.	mother	relationship: positive
I loved mother.	mother	None

Table 4.7: Examples of sub-projects from attributes

information the project contains. The realized language form of a turn can be generated when the project and turn type are known, so together, the project and the turn must contain all the information needed to build an inflected sentence. The turn type can be seen as a sentence template that has empty slots for parts of speech such as the subject, verb and object, which it receives from the project. Therefore, projects need to contain the subject, verb and object of the resulting sentence. In addition, the project contains information about time and the relative importance of the project. Importance is used to determine how strongly characters' moods are affected by the projects (see section 7.2).

A synthetic language [26] is a language that uses morphological inflections to convey relationships between words, as opposed to analytic languages, which mark relationships between words using word order and uninflecting marker words such as pre- and postpositions. Since the radio play system works in Finnish, a synthetic language, adding projects to turn types is not as simple as pasting a word in to an empty slot, like in analytic languages such as English. Instead, the insertion involves inflecting the incoming words according to grammatical rules.

Note that the field "object" is not used in the project context in the strict grammatical sense, but used loosely to mean the word that the action described by the verb is directed towards. In linguistic terms, the "object" corresponds to the semantic patient. An object type is also specified in the project to help the program decide what case to use when inflecting the object. The object type determines the kind of agent-patient relationship in question. It may describe the destination of movement, the object of affect, or simply the subject complement of a predicative sentence, such as in the sentence *I am your brother*.

This architecture, which separates the content of a sentence (project) from its form (turn type) enables the system to combine projects and turn types freely, resulting in a wide variety of sentences. Table 4.8 gives examples of projects, along with their realization in a simple turn type of the *statement* type. Table 4.9 illustrates turns from the radio play system formed using the following project:
subject: grandmother

Subject	Verb	Object	object type	time	realization
I	go	home	location	present	I go home.
I	go	home	location	past	I went home.
father	die	None	None	past	Father died.
weather	be	sunny	quality	present	The weather is sunny.

Table 4.8: Examples of projects

Action	Turn type group	Realization
inquire	statement	Do you know that grandmother has died?
inform	statement	Grandmother has died.
complain	surprise	Why didn't you tell me that grandmother has died?

Table 4.9: Different realization for the project "grandmother has died"

verb: die
object: none
time: perfect past

4.3 Story world

In order to maintain factual coherence, the radio play system must include a conceptualization of all the information that the system needs in order to create internally consistent narrative. Characters must be referred to with fixed names and described using non-conflicting terminology. The story should prevent the system from generating impossible or absurd sequences of events, such as characters dying, then talking normally in the next scene and objects appearing in different places without explanation.

In order to output consistent statements, the radio play generation system uses a single source of truth. This source is the story world. The story world consists of characters, objects and locations. Objects and locations have a name and a list of attributes, which help the characters form sentences when talking about them. In addition to concrete objects like vases, the story world contains objects that encapsulate abstract phenomena like the weather and relationships, enabling the characters to make consistent statements about these things as well.

Characters are the voices of the dialogue. All text in the scripts outputted by the system that will be spoken by actors and heard by listeners is speech by a character. In order for the dialogue to be vibrant and compelling, the characters' voices should paint an impression of the characters' individual personality. The characters should feel like unique personalities with motivations and quirks, not just different voices reading text. In addition, the speaking styles should differentiate the characters from each other to facilitate the listener's following of the dialogue.

On the other hand, characters are also decision-making agents within the storytelling system. All topics of conversation rise from the desires of the characters, and reactions to topics raised by another character should reflect the opinions of the character. These goals and preferences must also be consistent to be believable, but flexible in order to create change of direction, even twists in the plot, instead of becoming a monotonous back-and-forth argument.

Characters have names and distinct speaking styles. Each character has a set of beliefs, which is a (possibly imperfect) copy of the story world. Characters' beliefs may differ from each other, especially with regards to the attribution of story objects and locations. One character may consider a painting excellent, another bad. Characters also have a memory, a list of story events they know have happened. Memory determines which new events the character considers surprising. Characters also have affective states: personality and mood, which determine their behaviour as an agent. Personality is discussed in detail in section 5.1 and mood in section 5.2. For now it suffices to say that goals determine the structure and topics of what characters say, and affective states modify the format of the contents.

The story world also contains a causality knowledge base which all characters have access to. The knowledge base consists of rules that specify which events follow causally from other events. These rules are used to create a surprise project to form a surprise turn after hearing a surprising statement or proposal turn (table 4.10). The knowledge base includes a rule that states that sickness may result in death, so when character A is informed of character B's death, character A may express his surprise by forming the surprise turn "I didn't know that B had gotten sick". The causality rules are identical for all characters, although variation and new conflict could be added if characters had different assumptions on what the cause of the relative's death was - sickness, accident or murder.

	Subject	Verb	Object
Original	X	die	∅
Surprise	X	get sick	∅
Original	X	take	Y
Surprise	X	want	Y

Table 4.10: Causality rules are instructions for forming surprise projects.

5 Modeling affect

In order to generate radio play scripts that reflect the relationships and moods of its characters, the radio play generation system must have a representation of the relationships and moods. The representations should be dynamic and changeable, in order to allow changing emotions and moods, but also include some constants within an individual in order to express differences between personalities of characters.

The model of mood should capture the internal state of a character in a way that can be linked to different facets of the character's behaviour. Since mood is the variable in the radio play generation system that produces variance and color, the model of mood should be able to introduce variations in speech on a number of levels: lexical choices, length of turns and number of turns produced, for example.

In Gebhard's ALMA - A Layered Model of Affect [13], personality, mood and emotions are defined as three perspectives of the same phenomenon. All three aspects of affect represent an individual's orientation towards the outside world: they describe how the character is affected by outside affects and how they express their internal reality in their actions and speech. Personality is the slowest aspect to change, mood is medium-term and emotions change rapidly.

In the radio play generation system, personality includes the properties of a character that do not change during the story. Mood is the state of the character at a given point in time. Emotions are changes in this mood.

5.1 Personality

Characters' moods in the radio play are dynamic and may change from one extreme to another during the unfolding of the plot. However, to generate variance in generated scripts we need some starting parameters which can be randomized at the beginning of generation. These parameters, which determine the mood of a character at the start of the story, are called personality.

Developing numeric representations of personality is the main question of personality psychology. Numerous competing models rose and fell in the latter half of the 20th century,

with the Big Five model [11] coming out on top as a simple model with only five factors that robustly capture variations in human personality.

The Big Five model, also known as the OCEAN or CANOE model by the first letters of its components, factorizes the personality into five traits. These traits are openness (to experience), conscientiousness, extroversion, agreeableness and neuroticism. In the radio play generation system, each trait is represented by a floating point number in the range $[-1, 1]$.

5.2 Mood and emotions

Personality and emotion theories in the field of psychology may be rich and facilitate human understanding of behaviour, but they are not often applicable to simulated systems. A three-dimensional numeric model of mood was proposed by Mehrabian and Russell [18] in 1974. Two-dimensional models utilizing the pleasure and arousal axes are widely used and accepted [2] in psychology literature and research. Many emotional states such as happiness, sadness, excitement and boredom can be represented using the pleasure-arousal model (figure 5.1). The third dimension, dominance, has been less researched and used, although its relevance has been shown. The three-dimensional PAD space can be seen to model not only emotions but human experience in a broader sense [2]. In this view, the pleasure dimension represents feelings, the arousal dimension (activity of) thoughts and dominance represents outward behaviour.

The pleasure dimension is related to evaluation. It describes whether the agent is in a pleasant or unpleasant, favorable or unfavorable state. The arousal dimension relates to (physical and mental) activity and responsiveness. Extremely low arousal may indicate drowsiness or even sleep, whereas high arousal corresponds with frenzied excitement and high attentiveness. The dominance dimension, on the other hand, describes interpersonal phenomena such as control and influence. In the radio play generation system, dominance is closely related to the idea of *status*.

Status or status expression is a concept in dramaturgy and theatre that is used to describe power difference in the relationship between two characters. High status enables a character to influence others, whereas a character with low status will be influenced by others. Status expression is not equal to social status: a character with high social status character such as a king or CEO may express low status and a character with low social status such as a servant or cashier may express high status, although the two may often

Figure 5.1: Emotions placed on a pleasure-arousal field [2].

Pleasure	$0.21 \cdot E + 0.59 \cdot A + 0.19 \cdot N$
Arousal	$0.15 \cdot O + 0.30 \cdot A - 0.57 \cdot N$
Dominance	$0.25 \cdot O + 0.17 \cdot C + 0.60 \cdot E - 0.32 \cdot A$

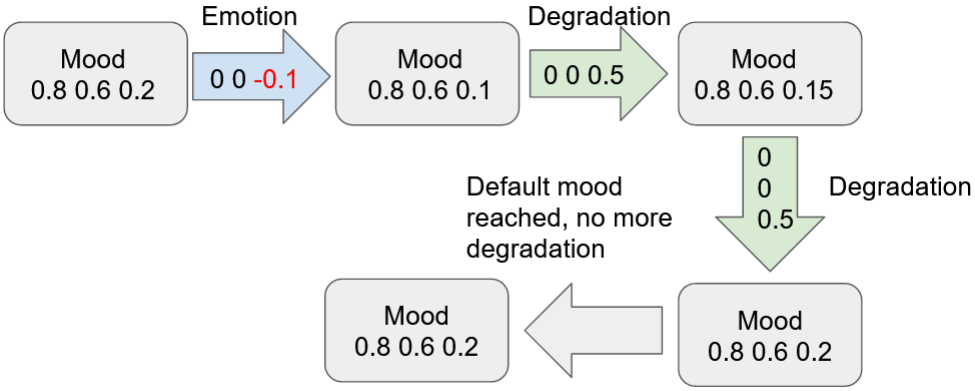
Table 5.1: Gebhard's [13] formulae for calculating default mood from Big Five personality traits: (O)penness, (C)onscientiousness, (E)xtroversion, (A)greeableness and (N)euroticism.

align.

In the radio play generation system, mood is a three dimensional PAD vector representing a character's affective state at a given time, and emotions are PAD vectors changing that state. Section 7.1 discusses how distinct emotions are defined and selected according to the situation the character is in.

Gebhard [13] defines a formula to compute default mood from a Big Five personality vector (table 5.1). The radio play generation system uses this formula to calculate default moods from personality traits, which are initialized randomly from a uniform distribution between -1 and 1. Personality does not directly affect characters' behaviour after default mood is calculated at the beginning of generation, except in one way: the neuroticism trait affects how much conflict the character can manage. This is explained in section 6.3.

Figure 5.2: Mood is affected by emotions and degradation.



In future development, it would be interesting to find more ways for personality to affect character behaviour. For example, the conscientiousness trait, also described as responsibility, could affect what kind of goals the character chooses: characters with high conscientiousness might select ambitious tasks whereas characters with low conscientiousness would be content with smaller goals. The openness trait, on the other hand, might affect how willing characters are to reveal their intentions directly instead of skirting around them.

If no events in the story affect a characters mood as time passes, the mood will degenerate towards the default mood. Figure 5.2 illustrates this process. The passing of time is measured as turns in the dialogue. Moods are reset to default at the beginning of a new scene, as time is assumed to have passed between scenes.

5.3 Relationships

Characters in a story are active agents and they have goals not only regards to events and objects in the story world, but goals regarding other characters' moods. Characters who don't like each other should be rude and harsh to each other, whereas characters who want to stay in good relations should be polite and gentle. In order to make decisions about how to influence the other's mood, character need to maintain representations of the other character's mood.

In the generation system, characters' relationships to one another are expressed as the emotional state (mood) that the character prefers the other to be in. When a character has

several possible ways to convey their intentions, they will select the one where the assumed emotional outcome brings the other character closest to the preferred state. However, a character only has a representation of the other's mood, and the representation may or may not be accurate.

6 Realization of emotional states

Modeling emotions in the way Gebhard [13] and Berov [4] have done is simple and produces a consistent, dynamic system. However, when generating dialogue, the goal is that the moods of characters are conveyed as impressions to the reader and do not remain only numbers in the system calculating them. In human speech, linguistic features play only a small part of detecting affect from speech [3], the more dominant features being acoustic, such as intensity and pitch. The goal in the radio generation system is to utilize such linguistic features to convey the emotional states to the reader.

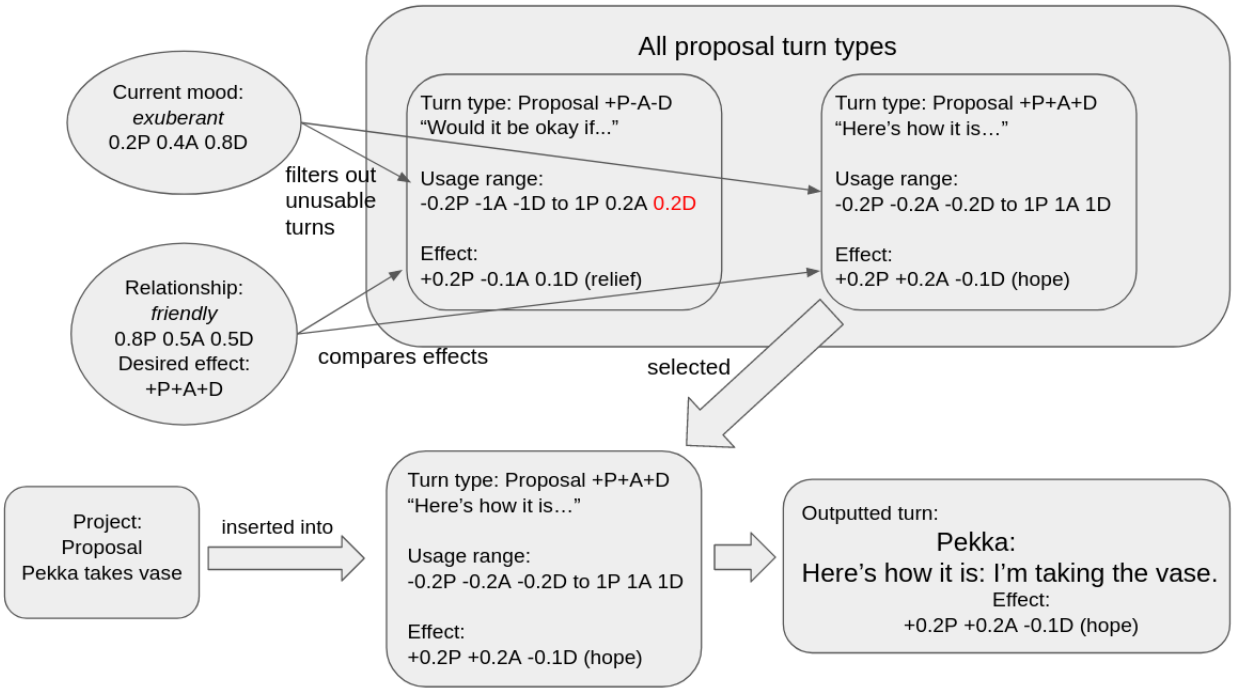
As we remember from figure 4.1, mood is both an influencer of turns taken by characters but is also influenced by turns heard. In this section, we inspect how mood influences turns on three different levels. First, we examine the mood expression system where turn types are selected according to a character's current mood (section 6.1). Second, we examine how mood is used to organize the flow of conversation on a higher level (section 6.2). Finally, we take a closer look at how affect is expressed in the Finnish language and how affective turn types can be written (section 6.4).

6.1 Mood-expressing turn types

In section 4.2.2, we discussed turn types, the abstraction of the form of a sentence that can be combined with a project - the content of a sentence - to form a turn, a speech act taken by a character. Turn types are grouped into first and second pair parts and further into functional categories: proposal, statement and surprise (first pair part) and refusal, acceptance and affirmation (second pair parts). We said before that turn types within a category should be equivalent information-wise, but didn't say much about why multiple ways to express the same intent were needed. That reason is to express the mood of the speaker.

As figure 4.6 roughly illustrated, turn type selection is determined by relationship and mood. Now that we have a model for mood and relationships, we can determine how these relate to specific turn types (figure 6.1). Both mood and relationship determine which turn type is selected for each turn. Turn types have a *usage range*, which relates to the speaker's current mood, and an *effect*, which is related to his relationship to the

Figure 6.1: Mood and relationship determine which turn type is selected to be combined with a project to form a turn.



listener.

A character's mood influences his selection of turn types by restricting turn type choice. This aims to reflect the everyday experience that word choice is not always deliberate, especially in extreme moods. In simple terms, a character may be too angry to use a friendly, polite turn type, or too happy to use a stern, demanding turn type.

Dividing the three-dimensional space of mood into eight sub-spaces, one for each combination of positive/negative mood components, the radio play system needs eight mood-expressing turn types for each category. Table 6.1 shows the turn types for the proposal category, filled with the project "I take the vase".

It is impossible to write turn types that exhibit the desired moods exactly. For one thing, language is ambiguous, and written language doubly so, lacking acoustic markers like intonation and stress. Even discounting differing interpretations, natural language is not continuously expressive in the sense that we can generate any number of expressions on a given scale. For example, if we define "Give me the vase right now or there will be consequences" as extremely dominant (1) and "Please can you let me have the vase" as extremely low dominance (-1), it is not reasonable to try to generate a range of expressions

+P+A+D	Here's how it is: I'm taking the vase.
+P+A-D	If you don't mind, I'll take the vase.
+P-A+D	I'd like to take the vase.
-P+A+D	I'm taking the vase, and that's final.
+P-A-D	Would it be okay if I took the vase?
-P+A-D	I would really like to take the vase.
-P-A+D	I'm taking the vase.
-P-A-D	I want to take the vase.

Table 6.1: The eight *proposal* turn types completed with vase taking project in the first person.

between the two at points -0.9, -0.8... etc.

To account for the fuzziness of the relation between natural language expression and numeric mood, turn types are written to roughly correspond to the eight quadrants of the mood space. Hence, it is not to be expected that opposite turn types (+P-A-D and -P-A-D being opposites on the P-axis, for example) should be symmetrical in the sense that their distance from the center of the mood space, i.e. their emotional magnitude, should be equal.

After turn type sentences have been written, they are evaluated on the PAD space and given a numeric value. Instead of corresponding to a precise point in the PAD space, each turn type is given two PAD values which define a range of moods. That particular turn type is usable if and only if a character's mood falls within that range. The turn type ranges should fill up the PAD space so that no point in the space is not covered by at least one turn type range. This way, when a character's mood changes, some turn types become unusable and others become usable, but in every mood there is at least one available turn type. Near the middle or origin of the space, many turn types will overlap, meaning that a character in a middle mood has usable many turn types.

Figure 6.2 illustrates turn type ranges in a two-dimensional mood space (the dominance factor is left out for the sake of simplicity). Turn type one is usable in emotional states ranging from -1 pleasure and -0.25 arousal to 0.33 pleasure and 1 arousal. Light grey areas in the corners indicate areas where only one turn type is usable, whereas in dark grey areas two turn types overlap. In the middle black area, all four turn types are usable. Note that the areas of each turn type are not symmetrical.

The example in figure 6.2 is not from the radio play generation system, as the third

Figure 6.2: Four turn types' ranges of use in the pleasure-arousal mood space.

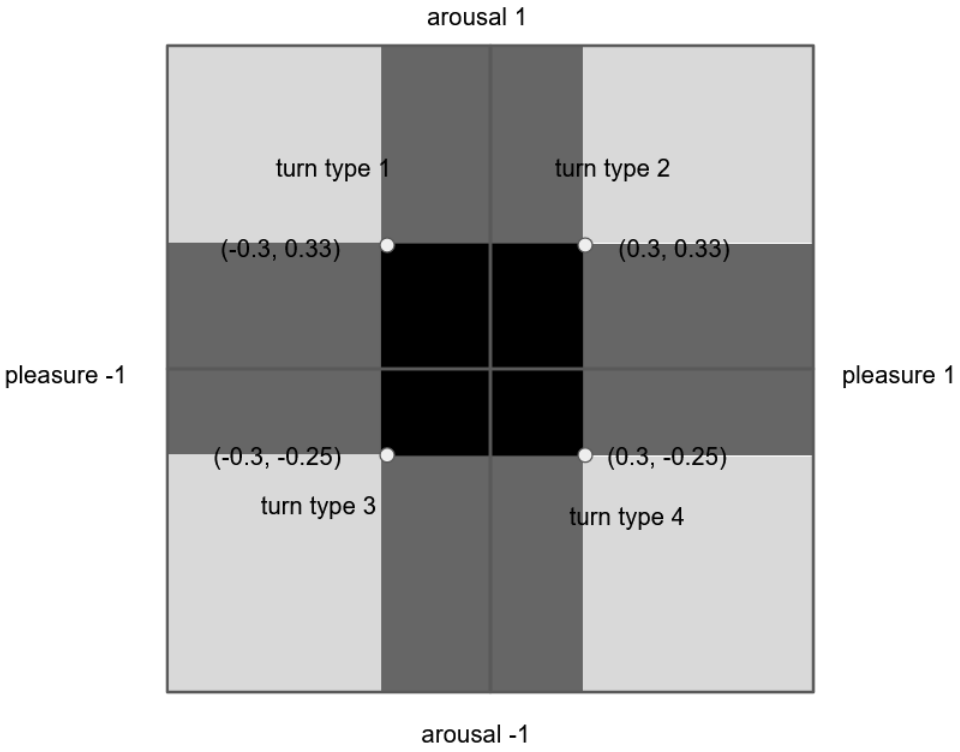
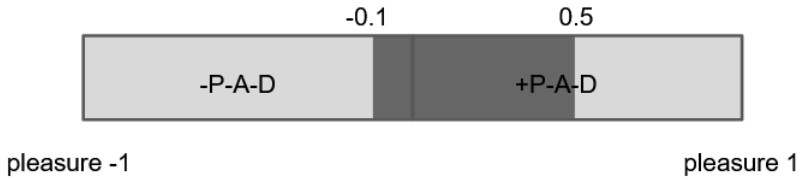


Figure 6.3: Four turn types' ranges of use in the pleasure-arousal mood space.



dimension, dominance, has been omitted for the sake of simplifying the visualization. The figure illustrates the main points of turn type use ranges. First, in the space of (-1, -1) to (1, 1) there is no unassigned white area, but all points in the area are painted with at least one turn type range. Second, the areas are slightly asymmetrical.

In the real radio play generation system, for example the two turns +P-A-D "Would it be okay if I took the vase?" and it's negative pleasure counterpart -P-A-D "I want to take the vase" are also asymmetrical (figure 6.3). "I want to take the vase" is quite neutral, so it can be used to up to +0.5 on the pleasure dimension. "Would it be okay if I took the vase", however, expresses more affect. It can only be used up -0.1 (dis)pleasure.

How are turn types written? Section 6.4 describes the way turn types are formed using knowledge about affective structures in language.

6.2 Mood and turn-taking

Generating dialogue using the hierarchical structure (figure 4.2) includes many situations where the system makes a choice. When generating sequences, the system must decide whether or not to include expansion sequences, and if so, what kind. After a sequence ends, the system needs to decide which character starts the next sequence.

One solution would be to make these choices stochastically. However, more cohesion can be built into the radio play script if choices follow deterministically from parameters. Since the radio play generation system aims to build dialogue that reflects its characters' moods, those moods are a logical choice for the basis of determining sequence and turn order.

Since dominance is the component of mood that relates to action and control, it makes sense for dominance to be the factor that governs turn-taking. Dominance determines which character speaks first. This invisible power struggle happens after a sequence is closed. The character with the higher dominance starts a new sequence, which may or may not include a pre-expansion. The pre-expansion is always made by the same character who speaks the main turn in the sequence. Remember that characters only start sequences if they have goals to advance, so if only one character has an unresolved goal, even a character with low dominance gets to start a sequence.

The arousal component of mood illustrates how lively and active a character is. High arousal increases the probability of expanding sequences with expansions. In other words, a character with high arousal is more prone to talking about topics in addition to the main topic. A character exhibiting low arousal, on the other hand, keeps to the point and expresses themselves minimally.

The pleasure component of mood doesn't affect whether characters say things, but rather *how* they say them. This reflects the view of Bakker et al. [2], which states that dominance relates to action, arousal to thought and pleasure to feelings. In the radio play generation system, dominance determines whether characters speak or not, arousal determines how long and complicated those speech acts (sequences) are by raising the probability of expansions, and pleasure (together with the other factors) dictates the contents of the bottom level speech acts, i.e. turns.

6.3 Internal conflict

According to Aristotle [6] one of ”the most powerful elements of emotional interest in Tragedy” is *peripeteia* or reversal of intention. This is the element of drama in which the course of action changes direction. The radio play generation system follows this rule of classical drama in order to bring each scene, and finally the whole story, to an end.

A scene in the story ends when all of its projects have been resolved. Projects are resolved in two ways. If the listening character outputs an accepting turn, the project is resolved as a success for the speaker. Alternatively, characters can abandon goals. Since characters are set to have different goals at the beginning of the story (otherwise the story would end very quickly), a change of goals must happen in order for the story to end.

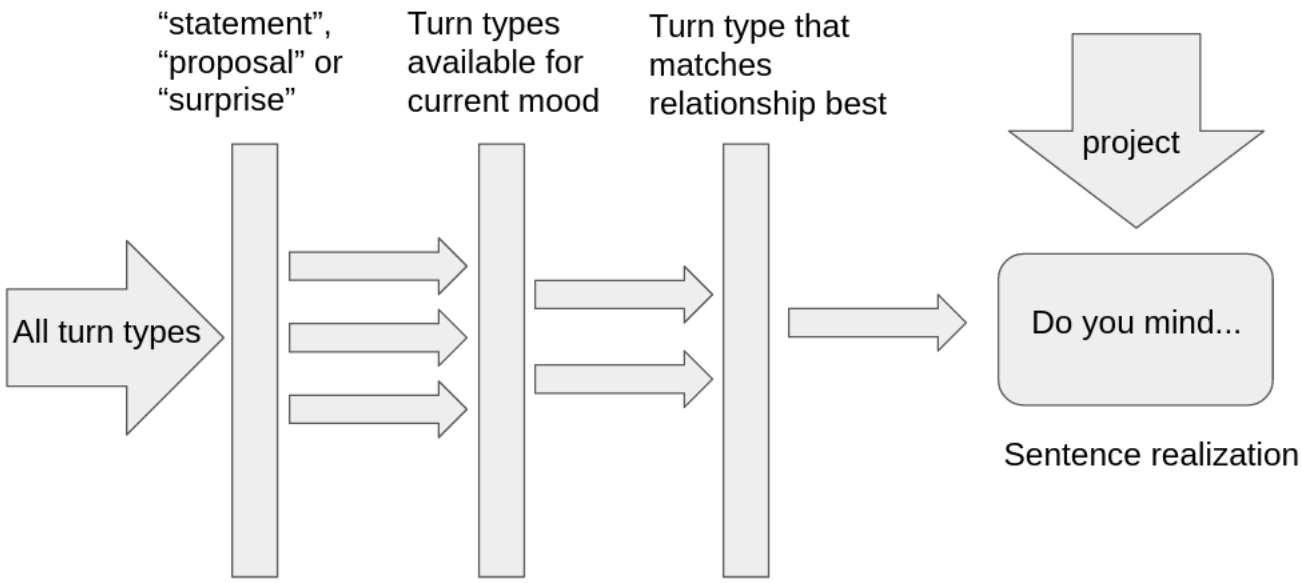
Abandonment of goals is caused by internal conflict. Internal conflict is a situation in which one character's goals are internally incongruous: the pursuit of one goal will hinder the other and vice versa. In the radio play generation system, this happens when a character's relationship with another character conflicts with pursuing his goal project. Remember that a relationship between characters is expressed as a mood state that the other character is preferred to exhibit. Therefore, the relationship can also be seen as a goal, a world state that the character wants to achieve, though it is not a sequence-generating goal.

Each turn, a character chooses an expression from a pool of turn types that differ in the way they affect the listening character's mood. When a character's own mood prevents him from using turns that would affect the listener in the way his relationship dictates, the character is internally conflicted.

Internal conflict is expressed in the outputted text by broken expressions: words in turns will be interlaced with three dots ”...”, conveying hesitation. Internal conflict causes stress, which characters have different capacities to tolerate, depending on their *neuroticism* personality trait. Stress tolerance is measured by the number of turns of internal conflict a character can withstand before being forced to resolve the conflict by either abandoning their goal or changing the relationship. In terms of the plot, this means that the character will either give up the valued inheritance object to maintain their relationship, or sacrifice the relationship in order to gain the inheritance object.

Rejecting statements and proposals by the other character also causes internal conflict, and will similarly result in change of goals and acceptance, if stress tolerance is surpassed. This ensures that the radio play will end even if the character pursuing their goal doesn't

Figure 6.4: Filtering process for an turn type in a turn.



experience internal conflict and consequently abandon their goal.

6.4 Affective language

In order to express characters' moods in the turn types they choose to use, we need to know how affect appears in the Finnish language. Characters will not be stating their mood explicitly by making statements like "I am angry". Instead, mood should be expressed subtly in the language characters use to communicate their goals.

Iso suomen kielioppi [14] (VISK, Finnish grammar handbook) describes affective constructions as a linguistic phenomenon. Affective constructions range from simple interjections to complex syntactic structures. The radio play generation system aims to employ as many as possible of these categories of linguistic expression of affect to generate vivid dialogue.

Interjections are short, non-inflecting exclamatory words such as *huh* and *wow*. As interjections don't affect the grammar of the rest of the sentence, they can be easily added to the beginning of many turn types. Affective interjections can be found in the surprise turns +P-A+D (Huh? Are you serious?) and -P+A+D (Huh? Why didn't you tell me that grandmother had gotten sick?).

Profanity is also an important way of conveying affect in language. Profanity can be seen

in the -P-A-D surprise turn *What the heck* and -P-A-D statement acceptance turn *Damn it*. Profanity seems a good fit for turns that express negative pleasure and negative arousal, as it is a way to express negative affect in a short turn, even in one word.

Affective syntactic structures are idiomatic ways to use grammatical structures to convey affect. VISK (§1713) recognizes affective interrogative expressions (also used in English) that employ interrogative words to express affect instead of their usual use to ask questions, such as *Miten kaunista olikaan, Mitä hittoa* (How beautiful it was, what the heck). The radio play uses this in the -P-A-D surprise turn *What the heck*.

The Finnish language (and also English) uses the imperative form to express affect in structures such as ”älä ihmeessä” (you don't say) and ”katos vaa” (well look at that). The radio play system utilizes this in the proposal refusal -P+A-D turn *Älä sano että haluat vaasin* (Don't say you want to take the vase) and the statement refusal +P-A+D turn *Älä höpsi* (Stop joking).

Another example of affective syntactic structures in Finnish is the infinitive clause. In this structure, a verb is used in infinitive form without a main verb. In the radio play's affective turn types, the affective infinitive is seen in the -P-A+D proposal acceptance turn ”Siitä vaan *ottamaan*” (go ahead and take it).

Iso suomen kielioppi [14] doesn't assess the quality of the aforementioned affective structures beyond positive/negative, only enumerates them as phenomena in the Finnish language. It remains the task of the writer of the radio play system to define how the affective structures translate to the PAD space. We assume that turn types with affective structures exhibit stronger affect and thus require more pronounced moods than turn types without.

Affect in language can also be examined on the lexical level, on the level of words and their affective connotations. One possible way to improve the radio play generation system could be to utilize affective words by swapping words in turns with affective synonyms or near-synonyms according to the speakers current mood. This replacement could be done also in turn types that don't have eight affective variations, like greetings.

7 Emotional effects of events

As depicted in figure 4.1, turn generation in the radio play system is a cycle where a character's mood affects the turns he outputs and hearing that turn modifies the listening character's mood, and so on. In section 6 we looked at how the three-dimensional PAD mood determines selection of turn types. Now we look at how the system determines what kind of effect hearing a turn has on the listener.

In order to act dynamically as agents in a story, characters' moods must not statically stay the same. Instead, characters should react to things the other character says. Emotional reactions or simply emotions are changes in mood, which can be expressed as a vector of length three, where the components are pleasure, mood and arousal. Different turns should have different effects, and the relationship between turn and effect should be believable.

The emotional reactions should enable the characters to potentially explore the whole three-dimensional mood space mapped by PAD (pleasure, arousal, dominance) values. Since there are eight affective variations for each turn type (one for each combination of negative/positive for each of the three components of mood), the goal of creating emotional reactions is to map these eight variations to a set of changes in mood that enable characters to move around the mood space.

The main problem that emotional reactions need to solve is realism. Reactions should be understandable in relation to the event that caused it, both in quality and in magnitude. To manage this, the radio play system must parameterize the effects of events and then map these parameterized events into PAD-valued emotions.

As illustrated in section 4.2, the radio play generation system conceptually splits each speech act or turn into its content (project) and form (turn type). Following this logic, the effects of heard turns are likewise split into two parts. First, there is the character's evaluation of the project it hears the other character express. This is discussed in section 7.2. Second, since projects can be expressed in many different ways, characters will also have an emotional reaction to the way it is expressed. For example, hearing a polite turn will please the listener, but a threatening turn may anger or frighten. The effects of turn types independent from the expressed project is discussed in section 7.3.

7.1 Categorizing emotions

Before going into the logic of translating a turn's project and turn type into an emotional effect, we will discuss conceptualizing emotions. Emotions are vectors that define a change in mood, for example the emotion (0, 0, -0.2) would change a mood of (0.8, 0.6, 0.7) to (0.8, 0.6, 0.5). But what phenomena do the numbers represent?

Steunebrink et al. [29] provide a way to model emotions based on a limited number of parameters. Figure 7.1 demonstrates the event evaluation decision tree. This model is parameterized by 5 parameters: type of event (consequence, action or aspect), valence of event (pleasant or unpleasant), cause of event (self or other), type of appraisal (prospective or actual) and familiarity, resulting in 30 different emotions.

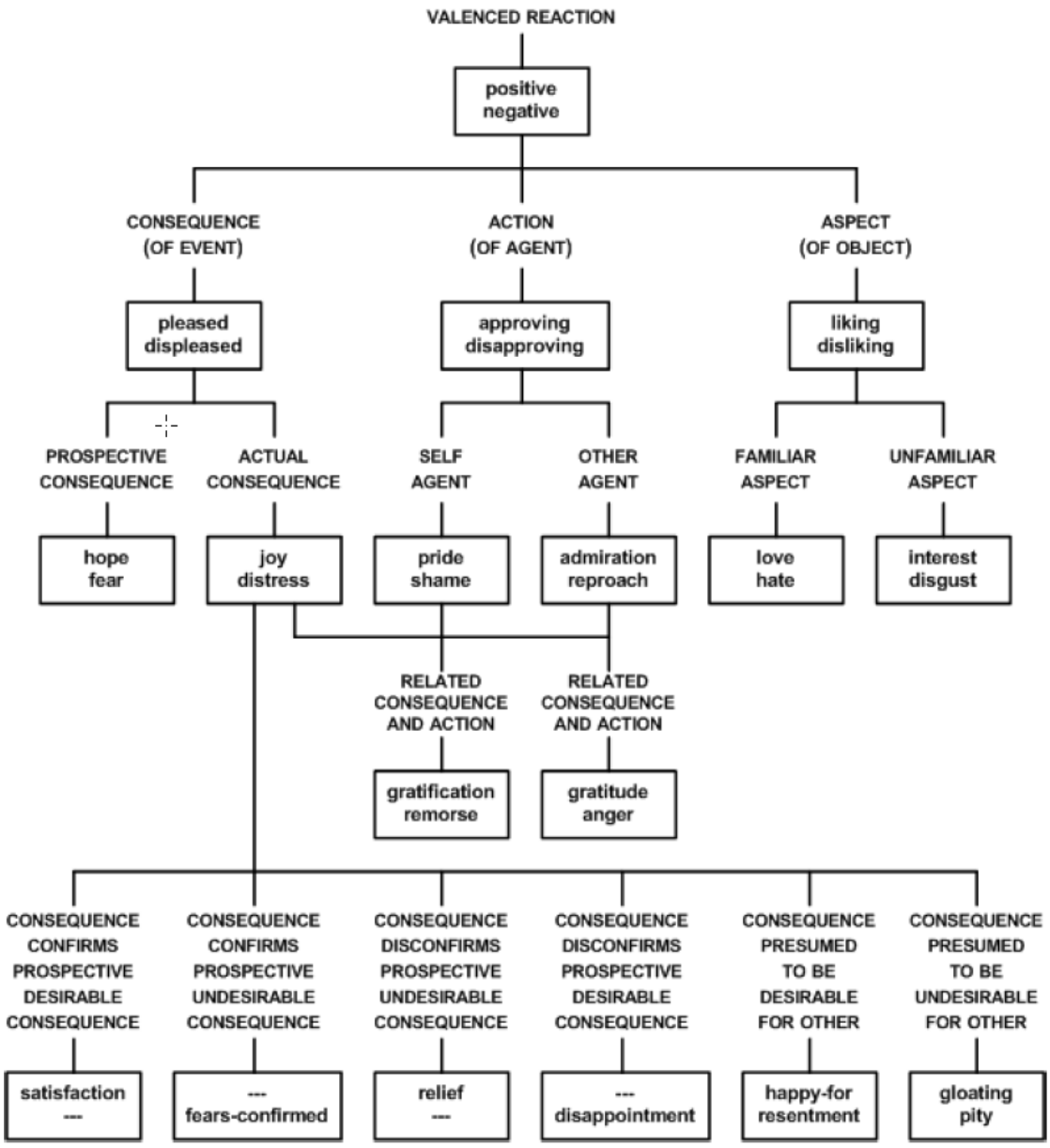
Parameterized affective evaluation is suitable for determining emotional affects of projects, as statement and proposal projects can be thought of as actions made by characters. Some of Steunebrink et al.'s parameters are not relevant in the radio play system, such as type of event and type of appraisal, as the radio play's project system does not model projects on that level of detail.

7.2 Effects of projects

In the radio play generation system, events are classified as good or bad. Characters also have model of possible events, which determines whether they are surprised by an incoming event. This gives the system three variables (surprised/not surprised, good/bad, caused by self/caused by other) which determine the class of emotion that will affect the character that hears the turn. Characters are always affected by turns spoken by the other character, not themselves. We use the term *speaker* to refer to the character expressing the first pair part in a sequence and the term *listener* for the character taking the second pair part (7.1). If there is a (surprise) insert expansion in the sequence, the roles are swapped: first pair part of the expansion is taken by the listener (of the sequence) and the second pair part by the speaker (of the sequence).

The opening turns (first pair parts) generated by the radio play system are divided into functional categories: statement, proposal and surprise. For statement turns we can define whether the statement is new information to the listener and whether it is pleasant or not. Using this reasoning, statements can be evaluated as provoking *joy* or *distress* if it is a

Figure 7.1: The decision tree for interpreting an event into an emotion by Steunebrink et al. [29].



Speaker	I'm taking the vase.
Listener	What?
Speaker	You heard me.
Listener	You are not taking the vase.

Table 7.1: Labeling the characters in a sequence as speaker and listener.

	Expected	**Surprise**
Favorable news	Satisfaction	Joy
Unfavorable news	Fears-confirmed	Distress

Table 7.2: Possible emotional reactions for statement first pair parts.

	Expected	**Surprise**
Acceptance	Satisfaction	Relief
Refusal	Fears-confirmed	Disappointment

Table 7.3: Possible emotional reactions for second pair parts.

surprise and *satisfaction* or *fears-confirmed* if expected. Table 7.2 displays this reasoning as a 2x2 matrix. Proposals are evaluated by the listener depending on whether the activity it offers is considered pleasant or not, resulting in *gratitude* or *anger*.

Surprise first pair parts always express a project that has already been introduced, since it is a reaction to news. Therefore, when evaluating an emotional response to hearing a surprise first pair part, it does not make sense to evaluate speaker's reaction to the surprise project, since it is a project originating from the speaker himself. Instead, the content evaluation of surprise first pair parts will neutral, and their emotional effect on the speaker will only depend on the turn's form, the turn type.

Similarly, second pair parts always echo the project of the first part they respond to, so evaluating the project from the point of view of the speaker is not meaningful. Instead, the speaker evaluates the turn by considering whether the second pair part is accepting or rejecting. Second pair parts are divided into two types: acceptance and rejection. The system does not generate dialogue where a speaker would make a proposal they want to see rejected, so accepting turns (made by the listener) will result in *satisfaction* or *relief*, if the acceptance was unexpected (a surprise). Table 7.3 illustrates the decision matrix for second pair parts.

Rejection turns evoke *fears-confirmed* or *disappointment*, depending on whether the rejection was expected or a surprise, respectively.

The OCC model [29] views emotions as changes in mood. Therefore emotions can be expressed using a three-dimensional floating point vector compromised of pleasure, arousal and dominance values, just like mood. The system uses numeric values provided by Steunebrink et al. (figure 7.2) to find representations for the different emotions evaluated

Figure 7.2: The OCC model [29] provides numeric values for emotions.

admiration	0.5	0.3	-0.2	+P+A-D	Dependent
anger	-0.51	0.59	0.25	-P+A+D	Hostile
disliking	-0.4	0.2	0.1	-P+A+D	Hostile
disappointment	-0.3	0.1	-0.4	-P+A-D	Anxious
distress	-0.4	-0.2	-0.5	-P-A-D	Bored
fear	-0.64	0.6	-0.43	-P+A-D	Anxious
fears_confirmed	-0.5	-0.3	-0.7	-P-A-D	Bored
gloating	0.3	-0.3	-0.1	+P-A-D	Docile
gratification	0.6	0.5	0.4	+P+A+D	Exuberant
gratitude	0.4	0.2	-0.3	+P+A-D	Dependent
happy_for	0.4	0.2	0.2	+P+A+D	Exuberant
hate	-0.6	0.6	0.3	-P+A+D	Hostile
hope	0.2	0.2	-0.1	+P+A-D	Dependent
joy	0.4	0.2	0.1	+P+A+D	Exuberant
liking	0.4	0.16	-0.24	+P+A-D	Dependent
love	0.3	0.1	0.2	+P+A+D	Exuberant
pity	-0.4	-0.2	-0.5	-P-A-D	Bored
pride	0.4	0.3	0.3	+P+A+D	Exuberant
relief	0.2	-0.3	0.4	+P-A+D	Relaxed
remorse	-0.3	0.1	-0.6	-P+A-D	Anxious
reproach	-0.3	-0.1	0.4	-P-A+D	Disdainful
resentment	-0.2	-0.3	-0.2	-P-A-D	Bored
satisfaction	0.3	-0.2	0.4	+P-A+D	Relaxed
shame	-0.3	0.1	-0.6	-P+A-D	Anxious

in the manner illustrated above.

Steunebrink et al.'s model doesn't account for differences between events of differing magnitude. Stubbing one's toe shouldn't elicit the same emotional response than having one's house burn down, although both are surprising, unpleasant events. To account for gradual differences between events, the radio play generation system includes a seriousness coefficient in order to express differing degrees of emotions. Projects carry a seriousness number from 0.1 to 1, 0.1 corresponding to trivial projects such as the weather and 1 to the most important projects such as relationships between characters. The main plot projects (informing the other character that a mutual relative has died, wanting to acquire the valuable inheritance object) carry an seriousness value of 1. Seriousness numbers are used as coefficients to modify the amplitude of the emotion being affected. For example, rejection to the project "the weather is nice" would result in *disappointment*, (-0.3 0.1 -0.4), but the seriousness coefficient 0.1 modifies this to (-0.3× 0.1 0.1× 0.1 -0.4× 0.1) = (-0.03 0.01 -0.04).

7.3 Effects of turn types

In addition to evaluation of projects, the radio play system must also be able to emotionally evaluate turn types, so that the same project realized as ”Could I please have the vase?” and ”I’m taking the vase and you have nothing to say about that” don’t have an identical effect.

Turn types take eight different emotionally flavoured variations, which together cover the whole three-dimensional PAD space from -1 to 1. The eight variations each govern a quadrant of the PAD cube, resulting in one variation for each combination of positive or negative pleasure, arousal and dominance. Without going into specific detail, can we say something about the signs of the effects of these turn types on the character who hears them? Does the turn lower or raise each component of mood?

One of the manifestations of empathy in humans is emotional contagion [28]. Through mimicry and mutual feedback, humans are affected by each others’ emotional states. The characters in the radio play will behave similarly with regards to pleasure and arousal.

Pleasure and arousal are *synergistic* mood components. What is meant by this is that using a turn that expresses pleasure begets pleasure in the listener also. A character expressing high arousal (being verbose and talkative) will encourage the listener to be more aroused as well. The corollary is also true: a turn expressing low pleasure lowers the listener’s pleasure level, and a character with low arousal will cause the other to slow down and minimize their speech also.

The same principle of contagion, however, does not work with dominance, since dominance in the radio play system’s model of mood, is a zero-sum game. Speaking and getting the other character to comply to one’s wishes are dominant actions and these actions are mutually exclusive. Both characters cannot speak at the same time, and both cannot change each other’s minds on the same issue. Therefore, dominance is *antagonistic*: expressing high dominance lowers the dominance of the listener. Conversely, hearing a turn expressing low dominance strengthens the dominance of the listener.

The rules of synergism and antagonism determine the directions of the emotion vectors resulting from turn types. The magnitude of each emotion remains to be decided individually for each turn type, taking into account the linguistic knowledge discussed in section 6.4. The conversion table shown in figure 7.2 serves as a starting point for estimating the effects of each turn type. For example, the -P+A-D turn ”Please don’t tell me you want

Usage quadrant	Effect
+P+A+D	+P+A-D
+P+A-D	+P+A+D
+P-A+D	+P-A-D
-P+A+D	-P+A-D
+P-A-D	+P-A+D
-P+A-D	-P+A+D
-P-A+D	-P-A-D
-P-A-D	-P-A+D

Table 7.4: Approximate effects of mood-expressing turn types using antagonistic/synergistic rules.

to take the vase" should, according to the rules of synergism and antagonism, produce a -P+A+D effect. In figure 7.2, such emotions are anger, disliking and hate. Hate and anger seem overly strong for reactions to the turn, so we will select *disliking*.

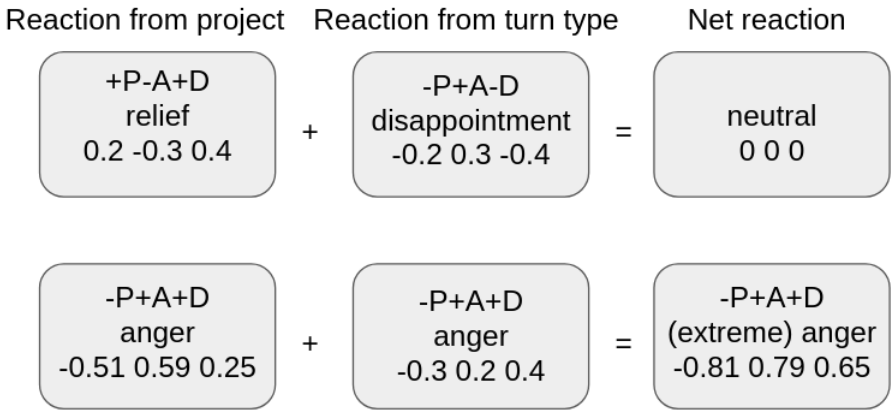
When choosing vector representations for turn type effects, it is important to remember to try to consider the form and wording of the turn regardless of whether it is an accepting turn or a rejection, since this information is already taken into account when calculating the emotional reaction to a project.

7.4 Combining emotional reactions

The emotional reactions to projects and turn types are calculated separately and combined via simple addition to form the final emotion vector. This emotion vector is added to the character's current mood to solve the new mood. Figure 7.3 illustrates two example cases of combining emotional reactions. In the first case, the reaction vectors for the project and turn type are exactly complimentary and cancel each other out. The resulting vector is the zero vector, in other words, no reaction.

In the second example, the two reaction vectors are of the same direction, roughly speaking. The combined vector is further from zero than either of its components. In other words, the emotional reactions amplify each other.

Figure 7.3: Combining project and turn type emotional reactions.



8 Experiments

In this section, we will examine the scripts generated by the radio play system built according to the architecture described in section 4 and mood model introduced in sections 5, 6 and 7. First, we will consider evaluation methods of natural language generation in section 8.1 to clearly define the perspectives we are using to examine the generated scripts. In section 8.2 we generate scripts using the predefined topics of grandmother's death in the first scene and taking the vase in the second scene. We evaluate the success of the scripts regarding the goals set for the system. In section 8.3, we generate scripts with different topics and examine how the principles of the system hold up when new projects are inserted. Scripts are generated in Finnish and translated to English for the purposes of this thesis. Original Finnish versions can be found in the appendices.

The goals of the radio play system can be divided into three points. First, we want to test if the proposed architecture of mood-expressing, interchangeable turn types manages to produce coherent dialogue within a constricted set of situations. Second, we assess whether the affective turn types succeed in communicating affect to the reader, and whether the communicated affect is of the intended quality. The third aspect of evaluation, assuming that character affect can be gathered from the text, is to consider whether the affective system creates dramatic tension.

8.1 Measures of quality

Evaluation of generated artifacts is a core question in natural language generation (NLG). The standard approach to automatic evaluation is to define a target text or ground truth and compare generated artifacts to this target, as recounted by Novikova et al. [19]. Methods of comparison differ from crude word-by-word overlap to more sophisticated methods like calculating semantic similarity. Methods which assume a single ground truth are more useful in NLG methods like machine translation and interactive dialogue generation, where a human produces half of the dialogue and the machine answers the human's prompts. However, Novikova et al. show that automatic evaluation methods fail to reflect human ratings. Automatic evaluation identifies low-quality artifacts, but fails particularly in the case of distinguishing mediocre from high quality.

At the time of writing this thesis, Adiwardana et al. at Google Research published a new state-of-the-art chatbot [1]. The evaluation of this work is done using a new metric called SSA, Sensibleness and Specificity Average. Adiwardana et al. argue that optimizing SSA leads to better language generators than optimizing other metrics such as perplexity [16]. As the name implies, SSA combines two qualities of natural language: sensibleness and specificity. Sensibleness requires that generated sentences are natural and grammatical and fit the context they are used in. Sensibleness is closely related to the concept of *coherence* used in this thesis.

However, according to Adiwardana et al., sensibleness alone is not sufficient for a good language generation system. Generated artifacts should also exhibit specificity. An example of a sensible but low-specificity sentence would be "I don't know", as the sentence can be coherently used in almost any context, but does not give much information or flavour to the text. If the NLG system doesn't optimize for specificity as well as sensibility, it will fall into the trap of picking safe but uninteresting options.

In section 8.2, as we examine examples of scripts generated by the radio play system, we will evaluate scripts both in terms of sensibility or coherence, looking out for unnatural or illogical sentences and upon finding them, investigate what shortcoming of the system may have caused them. In addition, we will consider the specificity of the generated artifacts. Are the turns taken by the characters interesting or bland? We can apply the concept of specificity to affective expression, since we are in particular interested in seeing if the turns seem to express specific emotional states, or if they could just as well be used in any emotional state.

8.2 Analysis of artifacts

8.2.1 A short script

The simplest scripts outputted by the radio play system look like the one in table 8.1. In the first scene, one character (Kalle) delivers the news about the relative's death and the other character Pekka simply replies to Kalle. All first pair parts are spoken by Kalle, except in the *surprise* insert expansion when Pekka expresses his surprise. Pekka first expresses disbelief once, but soon accepts the news as Kalle repeats it. Pekka's resistance only lasts for one turn, because he is high on the neuroticism trait.

In the second scene, both characters have the goal of acquiring the ornament, but Pekka

Scene 1		
0.18P 0.47A 0.44D	Kalle	Hi.
0.15P -0.43A 0.36D	Pekka	Hi.
0.18P 0.47A 0.44D	Kalle	I've got something to tell you. Grandma has died.
-0.55P -0.93A -0.44D	Pekka	What the heck?
-0.72P -0.23A -0.06D	Kalle	Yeah.
-0.55P -0.93A -0.44D	Pekka	This can't be true.
-0.72P -0.23A -0.06D	Kalle	Uh, you p-probably won't believe this, but grandma has died.
-1.00P -1.00A -0.94D	Pekka	Damn it.
Scene 2		
0.18P 0.47A 0.44D	Kalle	Can I take the vase?
0.75P -0.43A -0.04D	Pekka	Are you serious?
0.07P 0.66A 0.49D	Kalle	That's what I said.
0.75P -0.43A -0.04D	Pekka	You are not taking the vase.
0.07P 0.66A 0.49D	Kalle	Uh, c-can I... take the vase...?
1.00P -0.63A -0.54D	Pekka	Go ahead.

Table 8.1: Example of generated play that is minimal in length.

never gets to express his goal, as Kalle, having the higher dominance, always gets to speak first, with the exception of Pekka's surprise expansion ("Are you serious?").

In terms of coherence, the generated script is successful. There are no turns in the script that seem out of place or ungrammatical.

Both characters' default moods are similar, except for the arousal factor. In the first scene, hearing about grandma's death lowers Pekka's mood on all three dimensions. The same happens to Kalle as he hears Pekka's rude expression of surprise and rejection of the news. However, Kalle's simplistic turns don't express his mood, although they are not in conflict with it either. Pekka's short turns, which include profanity, serve better to express his low pleasure, low arousal mood. Pekka doesn't like the news: in fact, he seems like he is irritated by the whole conversation.

In the second scene, Kalle uses polite turns to ask for the vase, which raise Pekka's mood on the pleasure dimension but lowers it in terms of arousal and dominance. This mood of Pekka's could be described as lethargic but amused, or lazy. The turns Pekka uses (especially "Are you serious?" and "Go ahead") can be interpreted as fitting this mood, although they don't express the mood very strongly. Other interpretations would be possible as well. Kalle's mood doesn't change much from the default mood in scene two. He uses polite requests and the firm "That's what I said", which give the impression that he is in a happy but dominant mood. Kalle's turns are successful in terms of specificity, as they indicate that he is not angry or sad: they convey affect, though don't pinpoint it exactly.

The script cannot be said to display much dramatic tension, as the argument in the second scene is very short and the reader never finds out that Pekka wanted the vase too. Quite the opposite: the script gives the impression that Pekka doesn't care much about the vase at all, as even his refusal turn *you are not taking the vase* can be interpreted as amused disbelief instead of stern disagreement to the proposal. Perhaps Pekka considers the vase so ugly that he cannot understand why Kalle would even want it. Pekka's hesitation as he repeats his request does give the conversation's turning point some sense of dramatic weight, but it is short-lived, as the script ends immediately after.

8.2.2 Differing personality

To illustrate the effect the characters' personalities have on the content of the generated scripts, we will examine side by side two plays, where the only differing variable is one

character's personality. The other character's personality will remain the same in both scripts. Remember that given the start parameters (personalities and relationships), the generation of scripts is completely deterministic, involving no stochastic procedures.

Tables 8.2 and 8.3 illustrate a pair of scripts generated with parameters identical in all aspects except the second character Kalle's personality, which is close to neutral (0.2) on all traits in the first script, which gives him a default mood that is also close to neutral. In the second script, his personality is highly positive (0.9) on all traits, resulting in a default mood that is high in pleasure, neutral in arousal and quite high in dominance. The relationships of the characters are friendly, Pekka wants Kalle to be 0.8P 0.5A 0.5D and and Kalle prefers Pekka to be 0.9P 0.5A 0.5D.

Kalle delivers the news about grandma's death in the first scene politely, in accordance with his friendly relationship with Pekka, when his default mood is close to neutral in the first script. In the second script, however, Kalle's dominance level is too high for him to use the submissive turn "I don't want to disturb you...". Instead, the news is delivered in a more commanding tone: "I have something to tell you." As Kalle uses different turn types in the two scripts, they have different effects on Pekka, which causes Pekka to use different turn types to express his surprise.

In the second scene, a similar thing happens. In the first script, with a neutral personality and neutral default mood, Kalle uses a very polite turn type to propose that he should get grandma's vase. In the second script, Kalle is much more dominant, which clearly shows in his turn "Here's how it is...". Using such a dominating turn lowers Pekka's dominance so much that he doesn't get a turn to speak up about his own wish to have the vase. This makes the second script shorter in length than the first script, where Kalle's polite proposal gives Pekka confidence (raises his dominance) so that the next sequence is started by Pekka, who in turn proposes that he should take the vase. Ultimately, the vase still goes to Kalle in both scripts.

8.2.3 Differing relationship

Next, we leave personality parameters untouched and investigate how the relationship parameter affects the generation of scripts. We generate two scripts in table 8.4 and table 8.5, in both of which the characters' personalities are close to neutral (0.1 0.1 0.1 0.1 0.1 and 0.2 0.2 0.2 0.2 0.2). In the first script, Pekka is friendly towards Kalle (target mood 0.8P 0.5A 0.5D) but Kalle has something against Pekka (target mood -0.8P -0.5A

Mood	Speaker	Turn
Scene 1		
0.20P -0.02A 0.14D	Kalle	Hey.
0.10P -0.01A 0.07D	Pekka	Hello.
0.20P -0.02A 0.14D	Kalle	I don't want to disturb you but I should tell you that grandma has died.
-0.20P -0.21A -0.53D	Pekka	What the heck?
-0.70P -0.52A -0.36D	Kalle	Yeah.
-0.20P -0.21A -0.53D	Pekka	No way.
-0.70P -0.52A -0.36D	Kalle	Umm, g-grandma has died.
-1.00P -0.71A -1.00D	Pekka	Damn it.
Scene 2		
0.20P -0.02A 0.14D	Kalle	Hi.
0.10P -0.01A 0.07D	Pekka	Hello.
0.20P -0.02A 0.14D	Kalle	Would it be okay with you if I took the vase?
-0.21P 0.48A 0.42D	Pekka	Huh? Why haven't you told me that you want the vase?
-0.81P 1.00A 0.19D	Kalle	You heard what I said.
-0.21P 0.48A 0.42D	Pekka	You're not taking it.
-0.46P 0.73A 0.32D	Pekka	Uh, I'm taking the vase.
-1.00P 1.00A 0.34D	Kalle	Huh? Why haven't you told me that you want the vase?
-1.00P 1.00A 0.37D	Pekka	You heard what I said.
-1.00P 1.00A 0.34D	Kalle	You're not taking it.
-1.00P 1.00A 0.24D	Kalle	Umm, I'm... taking the vase and you d-don't have anything to say about t-that.
-1.00P 1.00A 0.52D	Pekka	Take it then, if it's so important to you.

Table 8.2: Script generated with personalities 0.1 0.1 0.1 0.1 0.1 and 0.2 0.2 0.2 0.2 0.2.

Mood	Speaker	Turn
0.89P -0.11A 0.63D	Kalle	Hello.
0.10P -0.01A 0.07D	Pekka	Hi.
0.89P -0.11A 0.63D	Kalle	I have something to tell you. Grandma has died.
-0.30P -0.11A -0.73D	Pekka	I didn't know grandma had gotten sick.
0.09P -0.11A 0.13D	Kalle	Unfortunately I didn't have time to tell you that grandma had gotten sick.
-0.30P -0.11A -0.73D	Pekka	I'm sure that's a lie.
0.09P -0.11A 0.13D	Kalle	I don't want to disturb you but I should tell you that grandma has died.
-0.40P -0.21A -1.00D	Pekka	Damn it.
Scene 2		
0.89P -0.11A 0.63D	Kalle	Hey.
0.10P -0.01A 0.07D	Pekka	Hi.
0.89P -0.11A 0.63D	Kalle	Here's how it is: I'm taking the vase.
-0.31P 0.78A 0.22D	Pekka	Huh? Why haven't you told me that you want the vase?
-0.12P 0.98A 0.68D	Kalle	Yeah. I wanted the vase last week.
-0.31P 0.78A 0.22D	Pekka	You're not taking it.
-0.12P 0.98A 0.68D	Kalle	Here's how it is: I'm taking the vase.
-0.62P 1.00A 0.27D	Pekka	Take it then, if it's so important to you.

Table 8.3: Script generated with personalities 0.1 0.1 0.1 0.1 0.1 and 0.9 0.9 0.9 0.9 0.9.

Mood	Speaker	Turn
Scene 1		
0.20P -0.02A 0.14D	Kalle	Hey.
0.10P -0.01A 0.07D	Pekka	Hi.
0.20P -0.02A 0.14D	Kalle	Grandma has died.
-0.60P -0.41A -0.53D	Pekka	What the heck?
-0.70P -0.52A -0.36D	Kalle	Yeah.
-0.60P -0.41A -0.53D	Pekka	This can't be.
-0.70P -0.52A -0.36D	Kalle	Grandma has died.
-1.00P -0.91A -1.00D	Pekka	Damn it.
Scene 2		
0.20P -0.02A 0.14D	Kalle	Hello.
0.10P -0.01A 0.07D	Pekka	Hi.
0.20P -0.02A 0.14D	Kalle	Umm, here's how it... i-is... I'm... t-taking the vase...
-0.31P 0.78A 0.22D	Pekka	Huh? Why haven't you told me that you want the vase?
-0.81P 1.00A 0.19D	Kalle	You heard what I said.
-0.31P 0.78A 0.22D	Pekka	You're not taking it.
-0.56P 1.00A 0.12D	Pekka	Uh, I'm taking the vase... and... you don't... have anything to s-say about that...
-1.00P 1.00A 0.34D	Kalle	Huh? Why haven't you told me that you want the vase?
-1.00P 1.00A 0.17D	Pekka	You heard what I said.
-1.00P 1.00A 0.34D	Kalle	Fine, take it then if it's so important to you.

Table 8.4: Script generated with asymmetrical relationship 0.8 0.5 0.5 and -0.8 -0.5 -1.

-1D). In the second script, both characters are equally hostile towards each other, wanting the other to suffer a mood of -0.8P -0.5A -1D.

The first scene plays out exactly the same in both scripts. Since Kalle's default mood and relationship are the same in both scripts, he chooses the same -P-A-D turn to inform Pekka of grandma's death. Pekka expresses his surprise using the same turn type in both scripts. In some situations, differing relationship could result in different turn types even though a character's mood is the same, but here it doesn't. Pekka's mood is below -0.2 on all dimensions, so he only has one turn type available to use: he is in a corner of the mood space.

In the second scene, Kalle with the higher dominance starts the first sequence, as seen

Mood	Speaker	Turn
Scene 1		
0.10P -0.01A 0.07D	Pekka	Hello.
0.20P -0.02A 0.14D	Kalle	Hello.
0.20P -0.02A 0.14D	Kalle	Grandma has died.
-0.60P -0.41A -0.53D	Pekka	What the heck?
-0.70P -0.52A -0.36D	Kalle	Yeah.
-0.60P -0.41A -0.53D	Pekka	This can't be.
-0.70P -0.52A -0.36D	Kalle	Grandma has died.
-1.00P -0.91A -1.00D	Pekka	Damn it.
Scene 2		
0.10P -0.01A 0.07D	Pekka	Hi.
0.20P -0.02A 0.14D	Kalle	Hey.
0.20P -0.02A 0.14D	Kalle	Umm... h-here's the thing: I'm t-taking... the v-vase.
-0.31P 0.78A 0.22D	Pekka	Huh ? Why haven't you told me that you want the vase?
-0.81P 1.00A 0.19D	Kalle	You heard what I said.
-0.31P 0.78A 0.22D	Pekka	You're not taking it.
-0.56P 1.00A 0.12D	Pekka	I'm taking the vase and you have nothing to say about that.
-1.00P 1.00A 0.34D	Kalle	Huh? Why haven't you told me that you want the vase?
-1.00P 1.00A 0.17D	Pekka	You heard what I said.
-1.00P 1.00A 0.34D	Kalle	Take it then, if it's so important to you.

Table 8.5: Script generated with relationships -0.8 -0.5 -1 in both directions.

in the scripts in the previous section as well. This scene also plays out almost exactly the same way. The only difference is that when Pekka states his intent to take the vase, he hesitates in the first script, since he wants Kalle to be in a good mood but his turn of choice (-P+A+D) lowers Kalle's pleasure and dominance levels. In the second script, when Pekka wants Kalle to be unhappy and submissive, he uses the same turn without hesitation. However, since Kalle hesitates in his proposal, the end result is the same in both scripts.

8.3 Different content

The radio play generation system separates the form of a sentence (the turn type) from its content (the project) for the purpose of providing a way to generate dialogue coloured by affect about any topic. The system has been built and tested using the following example projects:
Project type: statement
Subject: grandma
Verb: die
Object: none

and

Project type: proposal
Subject: [character]
Verb: take
Object: vase

Turn types have been written with the intention that they should express affect without restricting the content that can be inserted into them via projects. In this section, we will insert new projects into the system to test its flexibility.

If we replace the *statement* project with "inform other character about the birth of a baby" and the *proposal* project with "name the baby", the radio play system generates scripts such as the one depicted in table 8.6. In terms of coherence, the replacement works on many turns, but encounters some weird or illogical generations. These problems stem

Scene 1	
Pekka	Hello.
Pekka	A child has been born.
Kalle	I didn't know a child had been born.
Pekka	I'm sorry, a child was born last week.
Kalle	That is not true.
Pekka	A child has been born.
Kalle	I was afraid something like this was going to happen.
Scene 2	
Pekka	Hello.
Kalle	I'm naming the child and you have nothing to say about that.
Pekka	What the heck?
Kalle	Yeah.
Pekka	Don't tell me you want to name the child.
Pekka	I'm naming the child.
Kalle	What the heck?
Pekka	Yeah.
Kalle	Don't tell me you want to name the child.
Kalle	I'm naming the child.
Pekka	No! Not the child.
Kalle	I'm naming the child.
Pekka	Fine, name [him/her] then.

Table 8.6: Example of generated play with dead relative and inheritance projects swapped with birth and naming projects.

from a couple of reasons: one is that the structure of the system assumes that the news told by one character in scene one is bad. All turn types, including rejections, accepting turns and surprise turns are written with the assumption that the *statement*. This leads to absurd or at least unintentionally dark exchanges like "A child has been born" "I was afraid something like this was going to happen.".

To expand the scope of the stories created by the system, a mechanism should be added to control the nature of the *statement* project in the first screen: is it ill news or good news? This, of course, would also require more writing work, as each turn type in a turn type group should get a corresponding turn, expressing the same emotional state, but for

good news instead of bad.

In the radio plays generated so far, scene one always involves one character having a *statement* project and scene two involves both characters having *proposal* projects which are mutually incompatible. This was a choice made in the beginning of the process of building the system: a simple plot was chosen to test the affective turn type system. However, the system was built to be flexibly generative. Projects (goals) drive the generation of new sequences in the dialogue, but nothing in the system restricts the number of projects a character has in each scene, or how many scenes there are.

Table 8.7 illustrates a scene where both characters have two projects as their goals. Kalle has a *statement* project where he tells Pekka that he is going to move to the countryside, and a *proposal* project where he asks Pekka to move too. Pekka on the other hand has two *statement* projects: he tells Kalle that he has been laid off from his job and that he is going to sell his cabin. In this experiment, characters don't have conflicting goals, so no refusal turns are used. This means that there is no need for repeating the same project in multiple sequences to convince the other character.

In a longer conversation such as this, the amount of repetition of the same turn types becomes more prominent. Although there are eight affective turn types for each turn, in practice character's relationships seems to constrict the choices to only a few within one script. In this example, both characters even happen to choose the same turn types. It might be a good idea to modify the system so that it keeps track of used turn types and prefers, when possible, to pick a turn type that has not already been used in the text. On the other hand, repetition is a literary device that human authors intentionally use for stylistic effect and emphasis, so at least sometimes repetition in the radio play scripts may have a positive effect.

The sequences with different projects are generated independently of each other, but it's easy to interpret Pekka's two consecutive turns (10 and 11) as connected. If the script was acted and aired on radio, listeners would probably assume that Pekka's decision to accept Kalle's proposal (to come to the countryside with him) is related to the news Pekka reveals in the next sequence: that he has been laid off from his job. This creates a reasonably unified progression of sequences, however the conversation ends a bit abruptly.

On turn 7, Kalle hesitates because he knows that the turn type his mood forces him to use to express his proposal will have an effect on Pekka that conflicts with his relationship towards him. This hesitation is quite fortunate for dramatic tension. The listener can easily imagine dramatic reasons for the hesitation. The hesitation gives an impression that the

	Mood	Speaker	Turn
1	0.09P -0.50A 0.64D	Kalle	Hello.
2	0.10P 0.53A -0.27D	Pekka	Hi.
3	0.09P -0.50A 0.64D	Kalle	I have something to tell you. I'm moving to the countryside.
4	0.20P 0.13A 0.03D	Pekka	Are you serious?
5	0.49P -0.90A 0.44D	Kalle	That's what I said.
6	0.20P 0.13A 0.03D	Pekka	I was afraid something like this was going to happen.
7	0.49P -0.90A 0.44D	Kalle	Umm, c-could you c-come ... with me?
8	1.00P -0.07A -0.47D	Pekka	Are you serious?
9	0.38P -0.61A 0.49D	Kalle	That's what I said.
10	1.00P -0.07A -0.47D	Pekka	Sure, I'll come.
11	1.00P -0.27A -0.57D	Pekka	I was laid off.
12	0.48P -1.00A 0.79D	Kalle	Are you serious?
13	1.00P -0.67A -0.77D	Pekka	That's what I said.
14	0.48P -1.00A 0.79D	Kalle	I was afraid something like this was going to happen.
15	1.00P -0.67A -0.77D	Pekka	I've got some bad news. I'm selling the cabin.
16	-0.02P -1.00A -0.11D	Kalle	Are you serious?
17	1.00P -1.00A -0.97D	Pekka	Yeah.
18	-0.02P -1.00A -0.11D	Kalle	I was afraid something like this was going to happen.

Table 8.7: Generated scene where both characters have two projects.

question is important and the asker of the question is afraid of a rejecting answer. However, Pekka's surprise on turn 8 seems harsh in a sensitive interpersonal context. Pekka's sudden acceptance on turn 10 could even be construed as sarcastic in this interpretation.

9 Discussion and conclusions

The radio play system introduces a way to model characters' goals as conversation topics, projects, which are combinable with sentence forms called turn types to generate dialogue. The system models characters' mood in three dimensions: pleasure, arousal and dominance. Characters express their mood by selecting one of eight different affective turn types for each line of dialogue, resulting in conversations that differ in tone from passive-aggressive to polite. Dialogues starting from the same goals also differ in length and outcome in terms of whose goal is successful and who has to abandon his.

When generating language, a fundamentally cultural phenomenon, it does not make sense to only consider the data on its own. Language does not exist without humans to interpret it. The human mind construes meaning even in erratic text. Similarly to optical illusions, where the mind constructs whole shapes and patterns from incomplete data, because of top-down processes in information processing [17]. The brain has contextual models of what it predicts to find in a signal and the interpretation of stimuli is affected not only by the stimuli themselves (bottom-up processes) but what the brain expects to see. In the case of radio play scripts, a form of storytelling, a reader most likely expects the story to have a unified plot that has a beginning, turning point and resolution.

At the very least, a reader of dialogue will expect that the alternating turns will be related to each other; that the speakers hear each other's turns and respond to them. This may not actually be the case in computer-generated dialogue. In the radio play system, nested insert expansions create turns that don't have information about the turns that precede them chronologically. Table 9.1 shows a sequence with an insert expansion. The last line in the sequence is generated as a response to the first line with no information about the insert expansion in between. However, a human reader assumes that the response is a reaction to the line preceding it (I'm sorry, grandma got sick last week).

Pekka	Grandma has died.
Kalle	I didn't know grandma was sick.
Pekka	I'm sorry, grandma got sick last week.
Kalle	Well that's just great.

Table 9.1: Excerpt from script 8.3 (insert expansion indented for emphasis)

Human readers will make connections that are not intended by the program. Many turn types have many possible affective interpretations, but humans will most likely interpret coherence into a series of turns that may not have been generated coherently. This could have been utilized in the radio play system even more by writing turn types that are more vague and letting the coherence emerge from interpretation. Vague sentences such as ”I’ve made up my mind” and ”No! Don’t!” will be interpreted as connected, even if they were really just randomly selected by the computer.

Many turns in the scripts outputted by the radio play system seem sarcastic although the turn types are not intentionally written as sarcastic. This seems to happen when the affective quality that the turn type displays doesn’t seem to fit in its context.

Ultimately, the effect of a radio play is in the hands of the actors and director. The script can be seen as an intermediate phase of the finished product, the recorded radio play. Many elements in the generated scripts allow for creative choices by the director. For example, consecutive turns by the same character can be read together as one turn, or a pause can be left in between. Similarly, when the script generates hesitation, marked by three dots ”...” after a word, it is up to the director to decide how these are realized in enactment.

The use of prosodic features such as stress, pauses and pitch, which are absent from text but may be added in the enactment influence the ultimate form of the radio play. These features can be utilized to remedy the weaknesses of the generated scripts. For example the seeming abruptness of the change of heart can be helped by adding a longer pause between turns, and repetitive turns can be given variance by stressing them differently.

9.1 Modeling moods of others

The characters of the radio play system take other characters’ moods into account when making choices about what turns to use. They do this by directly accessing the world state, which contains all characters. In real life, we cannot read another people’s minds but have to imperfectly guess their internal state based on their behaviour. More realism and uncertainty could be added to the radio play generation system if characters in the radio play did not have direct access to the radio play’s representation of the other character’s state. Instead, characters would have a separate copy of the world model, which they update according to the speech acts they perform.

As turn types do not correspond to a singular point in the mood space, but a range of points forming the area of usage, characters cannot exactly pinpoint the mood of another by the turn type they use. Instead, they must make an approximation by taking the midpoint of the possible mood space. A character's estimation of another character's mood could be modeled using Bayesian maximization. Characters could hold the a priori assumption that others are in the same mood as themselves (assuming that a character has direct knowledge of his own mood). As speech acts are expressed, the estimation of the other character's mood would be updated according to what is the most likely mood for them to be in considering the speech acts so far.

Even more possibilities for misunderstanding rise if characters can have imperfect estimations of emotional reactions. This would mean that a character can have untrue information about the values of the parameters used for emotional reasoning. Recall from section 7.1 that emotional reaction depends on whether the project in the turn being reacted to is expected or a surprise, whether the activity proposed is pleasant or unpleasant. An imperfect model of the world could cause misinterpreting emotional reactions through these parameters. Here too, a human-like model might be that characters (sometimes mistakenly) assume that others will perceive activities in the same way as they themselves do.

In this phase of development, imperfect models of the moods of others were not implemented. With eight possible expressions in each situation, the affect system so simple that it probably couldn't express the nuances of misunderstood affective states. In future development, however, especially if new turn types that directly refer to mental states (for example: "Why are you so angry?") were added, modeling the characters' perceptions of other characters' moods would be interesting.

9.2 Turn types

In the radio play generation system, turn types are grouped into proposal, statement and surprise. The categorization is necessary in order to keep turn types combinable with each other and with projects. The experiments in section 8.3 show that the system works reasonably well even when the content of turns is changed, as long as the content lies within certain boundaries, ie. that the news in *statement* turns is considered unfavourable by both the bringer of news and the character who hears it.

In order to allow a wider range of topics to be discussed by characters in the radio play,

one of two solutions should be implemented. The first option is to rewrite turn types so that they don't specify the qualities of the topic but instead can be used with either good or bad news, either concrete or abstract subjects. The downside of this approach is that it limits affective expression. The alternative option is to separate turn type groups further. Instead of having three groups (proposal, statement and surprise) the system could have six groups for favourable proposal, unfavourable proposal, favourable statement, unfavourable statement, favourable surprise and unfavourable surprise. The downside of this is the need to write more turn types, especially if more parameterizations than one are needed. On the other hand, more turn types means more variety in the text, and many turn types can be reused in multiple groups - "Why didn't you tell me that ..." can be used to express surprise to either favourable or unfavourable news.

In longer generated scripts, like the one in table 8.7, the conversation becomes monotonous when it only consists of proposal, statement and surprise adjacency pairs. To bring more variety to the texts, and to enable a more complex plot, more turn type categories should be added. Categories such as the *question* to investigate projects further ("Why do you want the vase?", "When did grandma get sick?") and the *negotiation* to solve disagreements ("If you give me the vase, I'll be eternally grateful") could be among the next expansions to the system.

All *proposal* and *statement* turn types explicitly reference their project, whereas some *surprise* turn types do and some don't. As proposals and statements always start sequences and surprises occur as insert expansions, this means that the topic of conversation is always explicitly stated in the first turn of a sequence. It might be interesting to write some proposal and statement turn types that don't use project information (or don't use all of it) to create more uncertainty for the reader/listener. Table 9.2 illustrates possible ways to construct sequences with no project information in the first turn, the statement. The risk of having some project-agnostic turn types within a category is that if all the other turn types in the sequence also happen to be selected so that they don't display project information, the whole sequence becomes vague, as in the first example. Another problematic case is if a project-agnostic turn type is used in the statement turn but the surprise turn does display project information. As seen in example 4, this creates the unrealistic impression that the listener reads the speaker's thoughts or otherwise knows the news even though the speaker does not reveal it in the statement turn.

Examples 2 and 3, however, are fluent sequences. Arguably, they are better than the actual sequences generated by the system, where the project information is revealed on

1	No project information
Statement	It has happened.
Surprise	What? Are you serious?
Affirmation	You heard me.
Refusal	That can't be true.
2	Project information in last turn
Statement	I'm afraid I have some bad news.
Surprise	What? Are you serious?
Affirmation	You heard me.
Refusal	You must be mistaken, grandma can't be dead.
3	Project information in third turn
Statement	There's something I need to tell you.
Surprise	What? Are you serious?
Affirmation	I'm sorry, grandma died last week
Refusal	That can't be true.
4	Project information in second turn
Statement	It has happened
Surprise	Why didn't you tell me that grandma had gotten sick?
Affirmation	I'm sorry, grandma got sick last week
Refusal	That can't be true.

Table 9.2: Hypothetical sequences where project information revealed in different turns of the sequence.

the first turn. This creates a sense of tension or suspense in the first turns, when the reader doesn't yet know what has happened, but can infer from the affective expression of the characters that it is something unpleasant. Based on this experiment, it would be a good idea to add more project-agnostic turn types to the system to create a sense of dramatic suspense.

9.3 More topics

The project model gives the radio play system a way to separate the form of a turn from its content and give characters goals that they want to talk about in the dialogue. In this stage of development, projects are hand-selected and defined at the beginning of generation. However, in future development, it would be interesting to give the system more autonomy by letting it generate projects as well as dialogue. The simplest possibility would be to randomly pick nouns and verbs from a dictionary and build projects from them. To improve coherence, semantic dictionaries and language models should be used to select nouns and verbs that fit together in a project.

The next challenge for the radio play system is to see how well it can expand into longer stories, with more goals. In addition to having more pre-set goals at the start of a scene, characters could also acquire projects dynamically during the story, depending on what they hear. Hearing surprising news could create a goal related to that news, for example to find out more about the news. Hearing specific turn types or experiencing affective changes could also generate new goals: being insulted could make a character want to state their feelings about the insulting character, or becoming angry could create a project demanding the other to apologize.

One shortcoming of the current projects-as-goals system is that in its current form, it doesn't differentiate between goals that a character wants to realize by starting sequences and desires that reflect the character's passive preferences. This means that characters will accept all proposals unless they have a directly conflicting goal, such as wanting the same vase. Adding separate desires, projects that the character prefers to happen but doesn't actively pursue by starting sequences, would allow characters to reject proposals about projects without needing to propose the same thing themselves. This would bring the goal model on par with the BDI (beliefs, desires, intentions) model [23].

Acknowledgements

The author would like to thank Jessica Lindström and Otso Huopaniemi, who provided invaluable insights into conversation analysis and dramaturgy, respectively.

Bibliography

[1] D. Adiwardana, M.-T. Luong, D. R. So, J. Hall, N. Fiedel, R. Thoppilan, Z. Yang, A. Kulshreshtha, G. Nemade, Y. Lu, and Q. V. Le. *Towards a Human-like Open-Domain Chatbot*. Retrieved: April 10th 2020. Google, Palo Alto, United States, 2020. URL: https://arxiv.org/abs/2001.09977.

[2] I. Bakker, T. Van der Voordt, J. Boon, and P. Vink. *Pleasure, Arousal, Dominance: Mehrabian and Russell revisited. Current Psychology* 33, Springer, Heidelberg, Germany, pp. 405-421, Oct. 2014.

[3] A. Batliner, B. Schuller, D. Seppi, S. Steidl, L. Devillers, L. Vidrascu, T. Vogt, V. Aharonson, and N. Amir. *The Automatic Recognition of Emotions in Speech. Emotion-Oriented Systems*, Springer, Heidelberg, Germany, pp. 71-99, Jan. 2011.

[4] L. Berov. *Steering Plot Through Personality and Affect: An Extended BDI Model of Fictional Characters. KI 2017: Advances in Artificial Intelligence*, Springer International Publishing AG, Dortmund, Germany, pp. 293-299, Sept. 2017.

[5] L. Berov. *Towards a Computational Measure of Plot Tellability. 10th International Workshop on Intelligent Narrative Technologies*, AAAI, Snowbird Ski Resort, Utah, United States, pp. 169-175, Oct. 2017.

[6] S. Butcher. *Aristotle's Poetics*. New York City, United States: Macmillan and Co., 1902.

[7] J. E. Clark, S. Watson, and K. J. Friston. *What is mood? A computational perspective. Psychological Medicine* 48.14, Cambridge University Press, Cambridge, United States, pp. 2277–2284, Oct. 2018.

[8] S. Colton and G. Wiggins. *Computational creativity: The final frontier? ECAI'12: Proceedings of the 20th European Conference on Artificial Intelligence*, Frontiers in Artificial Intelligence and Applications, Amsterdam, The Netherlands, pp. 21-26, Jan. 2012.

[9] J. Cuddon. *The Penguin dictionary of literary terms and literary theory*. London, United Kingdom: Penguin Books, 1999, p. 877.

[10] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. Retrieved: April 10th 2020. Google, Palo Alto, United States, 2018. URL: https://arxiv.org/abs/1810.04805.

[11] J. M. Digman. *Personality Structure: Emergence of the Five-Factor Model. Annual Review of Psychology* 41.1, Annual Reviews, Palo Alto, United States, pp. 417-440, 1990.

[12] K. Friston. *The free-energy principle: a unified brain theory? Nature reviews. Neuroscience* 11, Nature Research, London, United Kingdom, pp. 127-138, Feb. 2010.

[13] P. Gebhard. *A Layered Model of Affect. 4th International Joint Conference of Autonomous Agents & Multi-Agent Systems (AAMAS'05)*, ACM Press, Utrecht, The Netherlands, pp. 29-36, Jan. 2005.

[14] A. Hakulinen, M. Vilkuna, R. Korhonen, V. Koivisto, T. R. Heinonen, and I. Alho. *Iso suomen kielioppi*. Suomalaisen Kirjallisuuden Seura, Helsinki, Finland, 2004, §1707–§1738. ISBN: 978-952-5446-35-7.

[15] O. Huopaniemi. Oral communication. Helsinki, Finland, 2019.

[16] F. Jelinek, R. L. Mercer, L. R. Bahl, and J. K. Baker. *Perplexity — a measure of the difficulty of speech recognition tasks. The Journal of the Acoustical Society of America* 62.S1, ASA, Chicago, United States, pp. S63, 1977.

[17] P. König, C. Chiang, and A. Stein. *Internal context and top-down processing. Behavioral and Brain Sciences* 20, Cambridge University Press, Cambridge, United States, pp. 691-692, Dec. 1997.

[18] A. Mehrabian. *Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in Temperament. Current Psychology* 14.4, Springer, Heidelberg, Germany, pp. 261-292, Dec. 1996.

[19] J. Novikova, O. Dušek, A. Cercas Curry, and V. Rieser. *Why We Need New Evaluation Metrics for NLG. Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, Copenhagen, Denmark, pp. 2241-2252, Jan. 2017.

[20] A. Ortony, G. Clore, and A. Collins. *The Cognitive Structure of Emotion. Contemporary Sociology* 18, Cambridge University Press, Cambridge, England, pp. 97-99, Jan. 1988.

[21] R. Pérez y Pérez. *Employing emotions to drive plot generation in a computer-based storyteller. Cognitive Systems Research* 8.2, Elsevier, Amsterdam, The Netherlands, pp. 89-109, 2007.

[22] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever. *Language Models are Unsupervised Multitask Learners*. Retrieved: April 10th, 2020. OpenAI, San Fransisco, United States, 2019. URL: https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf.

[23] A. Rao and M. Georgeff. *BDI agents: From theory to practice. Proceedings of the First International Conference on Multiagent Systems*, AAAI, San Francisco, United States, pp. 312-319, 1995.

[24] M. Rhodes. *An Analysis of Creativity. The Phi Delta Kappan* 42.7, Phi Delta Kappa International, Arlington, United States, pp. 305-310, 1961.

[25] J. D. Robinson. *Overall Structural Organization. The Handbook of Conversation Analysis*, editors T. Stivers and J. Sidnell, Wiley-Blackwell, Hoboken, United States, pp. 257-260, Nov. 2012.

[26] E. Sapir. *Types of Linguistic Structure. Language: An introduction to the study of speech*, Harcourt, Brace & World, New York, United States, pp. 120-146, 1929.

[27] E. Schegloff. *Sequence Organization in Interaction. A Primer in Conversation Analysis*, Cambridge University Press, Cambridge, United States, Jan. 2007.

[28] T. Singer and C. Lamm. *The Social Neuroscience of Empathy. Annals of the New York Academy of Sciences* 1156, Wiley-Blackwell, Hoboken, United States, pp. 81-96, Mar. 2009.

[29] B. R. Steunebrink, M. Dastani, and J.-J. C. Meyer. *The OCC Model Revisited*. Retrieved: April 10th, 2020. *Proceedings of the 4th Workshop on Emotion and Computing*, Jan. 2009. URL: https://www.researchgate.net/publication/228952254_The_OCC_model_revisited.

[30] I. Swartjes and M. Theune. *A Fabula Model for Emergent Narrative*. Ed. by S. Göbel, R. Malkewitz, and I. Iurgel. *Technologies for Interactive Digital Storytelling and Entertainment*, Springer Berlin Heidelberg, Darmstadt, Germany, pp. 49-60, 2006.

[31] G. Wiggins. *A preliminary framework for description, analysis and comparison of creative systems. Knowledge-Based Systems* 19, Elsevier, Amsterdam, The Netherlands, pp. 449-458, Nov. 2006.

Appendix A Generated radio-play scripts

Scene 1		
0.18P 0.47A 0.44D	Kalle	Moi.
0.15P -0.43A 0.36D	Pekka	Moi.
0.18P 0.47A 0.44D	Kalle	Minulla on kerrottavaa. Mummo on kuollut.
-0.55P -0.93A -0.44D	Pekka	Mitä hittoa?
-0.72P -0.23A -0.06D	Kalle	Niin.
-0.55P -0.93A -0.44D	Pekka	Ei ole totta.
-0.72P -0.23A -0.06D	Kalle	Tota, et v-varmaan usko tätä, mutta mummo on kuollut.
-1.00P -1.00A -0.94D	Pekka	Perkele.
Scene 2		
0.18P 0.47A 0.44D	Kalle	Voinko mä ottaa vaasin?
0.75P -0.43A -0.04D	Pekka	Oletko sä tosissasi?
0.07P 0.66A 0.49D	Kalle	Niinhän minä sanoin.
0.75P -0.43A -0.04D	Pekka	Sä et ota ruusua.
0.07P 0.66A 0.49D	Kalle	Tota, v-voinko mä... ottaa vaasin... ?
1.00P -0.63A -0.54D	Pekka	Ole hyvä.

Table A.1: Original outputted script for table 8.1.

Mood	Speaker	Turn
Scene 1		
0.10P -0.01A 0.07D	Pekka	Terve.
0.20P -0.02A 0.14D	Kalle	Moi.
0.20P -0.02A 0.14D	Kalle	En halua häiritä mutta minun pitäisi kertoa sinulle, että mummo on kuollut.
-0.20P -0.21A -0.53D	Pekka	Mitä hittoa?
-0.70P -0.52A -0.36D	Kalle	Niin.
-0.20P -0.21A -0.53D	Pekka	Ei ole totta.
-0.70P -0.52A -0.36D	Kalle	Tuota, m-mummo on kuollut.
-1.00P -0.71A -1.00D	Pekka	Ei helvetti.
Scene 2		
0.10P -0.01A 0.07D	Pekka	Terve.
0.20P -0.02A 0.14D	Kalle	Moi.
0.20P -0.02A 0.14D	Kalle	Sopisiko sinulle mitenkään että otan vaasin?
-0.21P 0.48A 0.42D	Pekka	Täh? Mikset ole kertonut minulle että sä olet halunnut vaasin?
-0.81P 1.00A 0.19D	Kalle	Kuulit kyllä mitä sanoin.
-0.21P 0.48A 0.42D	Pekka	Sä et ota.
-0.46P 0.73A 0.32D	Pekka	Öö, mä otan vaasin.
-1.00P 1.00A 0.34D	Kalle	Täh? Mikset ole kertonut minulle että olet halunnut vaasin?
-1.00P 1.00A 0.37D	Pekka	Kuulit kyllä mitä sanoin.
-1.00P 1.00A 0.34D	Kalle	Et ota.
-1.00P 1.00A 0.24D	Kalle	Tuota, otan ... vaasin ja sinulla e-ei o-ole siihen mitään s-sanomista.
-1.00P 1.00A 0.52D	Pekka	Ota sitten jos se on niin tärkeä sinulle.

Table A.2: Original generated script for table 8.2.

Mood	Speaker	Turn
0.10P -0.01A 0.07D	Pekka	Hei.
0.89P -0.11A 0.63D	Kalle	Terve.
0.89P -0.11A 0.63D	Kalle	Minulla on kerrottavaa. Mummo on kuollut.
-0.30P -0.11A -0.73D	Pekka	En tiennyt, että mummo on sairastunut.
0.09P -0.11A 0.13D	Kalle	Valitettavasti en ehtinyt ilmoittaa sinulle, että mummo on sairastunut.
-0.30P -0.11A -0.73D	Pekka	Ei muuten pidä paikkaansa.
0.09P -0.11A 0.13D	Kalle	En halua häiritä, mutta minun pitäisi kertoa sinulle, että mummo on kuollut.
-0.40P -0.21A -1.00D	Pekka	Perkele.
Scene 2		
0.10P -0.01A 0.07D	Pekka	Hei.
0.89P -0.11A 0.63D	Kalle	Moi.
0.89P -0.11A 0.63D	Kalle	Nyt on sillä tavalla, että minä otan vaasin.
-0.31P 0.78A 0.22D	Pekka	Täh? Mikset ole kertonut minulle, että sä olet halunnut vaasin?
-0.12P 0.98A 0.68D	Kalle	Niin. Minä halusin vaasin viime viikolla.
-0.31P 0.78A 0.22D	Pekka	Sä et ota.
-0.12P 0.98A 0.68D	Kalle	Nyt on niin, että minä otan vaasin.
-0.62P 1.00A 0.27D	Pekka	Ota sitten, jos se on niin tärkeä sinulle.

Table A.3: Original generated script for table 8.3.

Mood	Speaker	Turn
Scene 1		
0.10P -0.01A 0.07D	Pekka	Moi.
0.20P -0.02A 0.14D	Kalle	Hei.
0.20P -0.02A 0.14D	Kalle	Mummo on kuollut.
-0.60P -0.41A -0.53D	Pekka	Mitä hittoa?
-0.70P -0.52A -0.36D	Kalle	Niin.
-0.60P -0.41A -0.53D	Pekka	Ei ole totta.
-0.70P -0.52A -0.36D	Kalle	Mummo on kuollut.
-1.00P -0.91A -1.00D	Pekka	Ei helvetti.
Scene 2		
0.10P -0.01A 0.07D	Pekka	Hei.
0.20P -0.02A 0.14D	Kalle	Terve.
0.20P -0.02A 0.14D	Kalle	Tota, nyt on... niin e-että... mä... o-otan vaasin...
-0.31P 0.78A 0.22D	Pekka	Täh? Mikset ole kertonut minulle että olet halunnut vaasin?
-0.81P 1.00A 0.19D	Kalle	Kuulit kyllä mitä sanoin.
-0.31P 0.78A 0.22D	Pekka	Et ota.
-0.56P 1.00A 0.12D	Pekka	Tuota, otan vaasin... ja... sinulla ei... ole siihen m-mitään sanomista...
-1.00P 1.00A 0.34D	Kalle	Täh? Mikset ole kertonut minulle että sä olet halunnut vaasin?
-1.00P 1.00A 0.17D	Pekka	Kuulit kyllä mitä sanoin.
-1.00P 1.00A 0.34D	Kalle	No ota sitten, jos se on niin tärkeä sinulle.

Table A.4: Original generated script for table 8.4.

Mood	Speaker	Turn
Scene 1		
0.10P -0.01A 0.07D	Pekka	Terve.
0.20P -0.02A 0.14D	Kalle	Terve.
0.20P -0.02A 0.14D	Kalle	Mummo on kuollut.
-0.60P -0.41A -0.53D	Pekka	Mitä hittoa?
-0.70P -0.52A -0.36D	Kalle	Niin.
-0.60P -0.41A -0.53D	Pekka	Ei ole totta.
-0.70P -0.52A -0.36D	Kalle	Mummo on kuollut.
-1.00P -0.91A -1.00D	Pekka	Helvetti.
Scene 2		
0.10P -0.01A 0.07D	Pekka	Hei.
0.20P -0.02A 0.14D	Kalle	Moi.
0.20P -0.02A 0.14D	Kalle	Öö... n-nyt on sillä tavalla, että m-mä otan ... v-vaasin.
-0.31P 0.78A 0.22D	Pekka	Täh? Mikset ole kertonut minulle että olet halunnut vaasin?
-0.81P 1.00A 0.19D	Kalle	Kuulit kyllä mitä sanoin.
-0.31P 0.78A 0.22D	Pekka	Et ota.
-0.56P 1.00A 0.12D	Pekka	Otan vaasin ja sinulla ei ole siihen mitään sanomista.
-1.00P 1.00A 0.34D	Kalle	Täh? Mikset ole kertonut minulle että sä olet halunnut vaasin?
-1.00P 1.00A 0.17D	Pekka	Kuulit kyllä mitä sanoin.
-1.00P 1.00A 0.34D	Kalle	No ota sitten, jos se on niin tärkeä sinulle.

Table A.5: Original generated script for table 8.5.

Speaker	Turn
Scene 1	
Pekka	Terve.
Pekka	Minulla on uutisia. Lapsi on syntynyt.
Kalle	En tiennyt että lapsi on syntynyt.
Pekka	Olen pahoillani, lapsi syntyi viime viikolla.
Kalle	Ei muuten pidä paikkaansa.
Pekka	Lapsi on syntynyt.
Kalle	Se tästä vielä puuttuikin.
Scene 2	
Pekka	Terve.
Kalle	Mä nimeän lapsen ja sulla ei ole siihen mitään sanomista.
Pekka	Mitä hittoa?
Kalle	Niin.
Pekka	Älä vain sano että sä haluat nimetä lapsen.
Pekka	Mä nimeän lapsen.
Kalle	Mitä hittoa?
Pekka	Niin.
Kalle	Älä vain sano että sä haluat nimetä lapsen.
Kalle	Mä nimeän lapsen.
Pekka	Ei! Ei lasta.
Kalle	Mä nimeän lapsen.
Pekka	No nimeä sitten.

Table A.6: Original generated script for table 8.6.

Mood Scene 1	Speaker	Turn
0.09P -0.50A 0.64D	Kalle	Terve.
0.10P 0.53A -0.27D	Pekka	Moi.
0.09P -0.50A 0.64D	Kalle	Muutan maalle.
0.20P 0.13A 0.03D	Pekka	Oletko tosissasi?
0.49P -0.90A 0.44D	Kalle	Niinhän minä sanoin.
0.20P 0.13A 0.03D	Pekka	Pelkäsinkin, että jotain tällaista tapahtuu.
0.49P -0.90A 0.44D	Kalle	Tuota, v-voisitko l-lähteä ... mukaan?
1.00P -0.07A -0.47D	Pekka	Oletko tosissasi?
0.38P -0.61A 0.49D	Kalle	Niinhän minä sanoin.
1.00P -0.07A -0.47D	Pekka	Lähden toki.
1.00P -0.27A -0.57D	Pekka	Sain potkut.
0.48P -1.00A 0.79D	Kalle	Oletko tosissasi?
1.00P -0.67A -0.77D	Pekka	Niinhän minä sanoin.
0.48P -1.00A 0.79D	Kalle	Pelkäsinkin, että jotain tällaista tapahtuu.
1.00P -0.67A -0.77D	Pekka	Myyn mökin.
-0.02P -1.00A -0.11D	Kalle	Oletko tosissasi?
1.00P -1.00A -0.97D	Pekka	Niin.
-0.02P -1.00A -0.11D	Kalle	Pelkäsinkin, että jotain tällaista tapahtuu.

Table A.7: Original generated text for the script in table 8.7